second edition

CMOS DIGITAL INTEGRATED CIRCUITS

Analysis and Design

SUNG-MO (STEVE) KANG
University of Illinois at Urbana- Champaign

YUSUF LEBLEBICI
Worcester Polytechnic Institute
Swiss Federal Institute of Technology-Lausanne

WCB McGraw-Hill

Boston Burr Ridge, IL Dubuque, IA Madison, WI New York San Francisco St. Louis
Bangkok Bogotá Caracas Lisbon London Madrid
Mexico City Milan New Delhi Seoul Singapore Sydney Taipei Toronto

WCB/McGraw-Hill

A Division of The McGraw-Hill Companies

To Myoung-A, Jennifer and Jefferey,
and Anıl and Ebru

CMOS DIGITAL INTEGRATED CIRCUITS: ANALYSIS AND DESIGN

This book is printed on acid-free paper.

1 2 3 4 5 6 7 8 9 0 DOC/DOC 9 3 2 1 0 9 8

ISBN 0-07-292507-8

Vice president and editorial director: Kevin T. Kane
Publisher: Thomas Casson
Executive editor: Elizabeth A. Jones
Senior developmental editor: Kelley Butcher
Marketing manager: John T. Wannemacher
Project manager: Karen Nelson
Senior production supervisor: Heather D. Burbridge
Senior designer: Lorna Lo
Cover designer: Nicole Leong
Supplement coordinator: Linda Huenecke
Typeface: 10/12 Times Roman
Printer: R. R. Donnelley & Sons Company

Library of Congress Cataloging-in-Publication Data

Kang, Sung Mo, (date)
CMOS digital integrated circuits: analysis and design/Sung-Mo (Steve) Kang, Yusuf Leblebici.
–2nd ed.
p. cm.
Includes index.
ISBN 0-07-292507-8
1. Metal oxide semiconductors, Complementary. 2. Digital integrated circuits. I. Leblebici, Yusuf.
II. Title.
TK7871.99M44K36 1999
621.39'5--dc21

98-28075

http://www.mhhe.com

Next, substitute V_{out} in the KCL equation (5.42), to obtain

$$2\cdot\left[2\cdot(V_{IH}-1)\cdot(0.5\,V_{IH}-0.55)-(0.5\,V_{IH}-0.55)^2\right]=\frac{1}{3}\cdot(2.95)^2$$

The solution of this simple quadratic equation yields two values for V_{IH} .

$$V_{IH}=\begin{cases}-0.35\ \text{V}\\ \underline{\underline{2.43\ \text{V}}}\end{cases}$$

where $V_{IH}=2.43$ V is the physically correct solution. The output voltage level at this point is calculated as

$$V_{out}=0.5\cdot 2.43-0.55=0.67\ \text{V}$$

With this updated output voltage value, we can now reevaluate the load threshold voltage as $V_{T,load}(V_{out}=0.67\ \text{V})=-2.9$ V, and the $(dV_{T,load}/V_{out})$ value as

$$\frac{dV_{T,load}}{dV_{out}}=0.18$$

Note both of these values are fairly close to those used at the beginning of this iteration process. Repeating the iterative calculation will provide only a marginal improvement of accuracy, thus, we may accept $V_{IH}=2.43$ V as a good estimate.

In conclusion, the noise margins for high signal levels and for low signal levels can be found as follows:

$$NM_H=V_{OH}-V_{IH}=2.57\ \text{V}$$
$$NM_L=V_{IL}-V_{OL}=1.17\ \text{V}$$

5.4. CMOS Inverter

All of the inverter circuits considered so far had the general circuit structure shown in Fig. 5.3, consisting of an enhancement-type nMOS driver transistor and a load device which can be a resistor, an enhancement-type nMOS transistor, or a depletion-type nMOS transistor acting as a nonlinear resistor. In this general configuration, the input signal is always applied to the gate of the driver transistor, and the operation of the inverter is controlled primarily by switching the driver. Now, we will turn our attention to a radically different inverter structure, which consists of an enhancement-type nMOS transistor and an enhancement-type pMOS transistor, operating in complementary mode (Fig. 5.16).

This configuration is called Complementary MOS (CMOS). The circuit topology is complementary push-pull in the sense that for high input, the nMOS transistor drives (pulls down) the output node while the pMOS transistor acts as the load, and for low input the pMOS transistor drives (pulls up) the output node while the nMOS transistor acts as the load. Consequently, both devices contribute equally to the circuit operation characteristics.

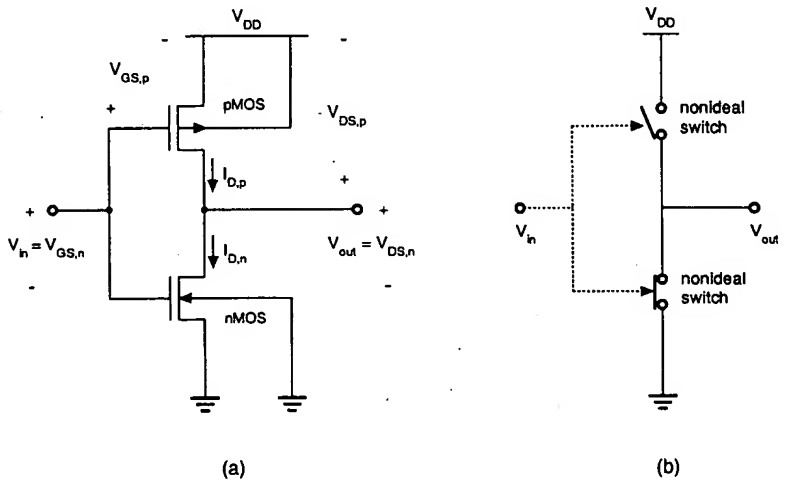


Figure 5.16. (a) CMOS inverter circuit. (b) Simplified view of the CMOS inverter, consisting of two complementary nonideal switches.

The CMOS inverter has two important advantages over the other inverter configurations. The first and perhaps the most important advantage is that the steady-state power dissipation of the CMOS inverter circuit is virtually negligible, except for small power dissipation due to leakage currents. In all other inverter structures examined so far, a nonzero steady-state current is drawn from the power source when the driver transistor is turned on, which results in a significant DC power consumption. The other advantages of the CMOS configuration are that the voltage transfer characteristic (VTC) exhibits a full output voltage swing between 0 V and V_{DD}, and that the VTC transition is usually very sharp. Thus, the VTC of the CMOS inverter resembles that of an ideal inverter.

Since nMOS and pMOS transistors must be fabricated on the same chip side- by-side, the CMOS process is more complex than the standard nMOS-only process. In particular, the CMOS process must provide an n-type substrate for the pMOS transistors and a p-type substrate for the nMOS transistors. This can be achieved by building either n-type *tubs* (wells) on a p-type wafer, or by building p-type tubs on an n-type wafer (cf. Chapter 2). In addition, the close proximity of an nMOS and a pMOS transistor may lead to the formation of two parasitic bipolar transistors, causing a *latch-up* condition. In order to prevent this undesirable effect, additional *guard rings* must be built around the nMOS and the pMOS transistors as well (cf. Chapter 13). The increased process complexity of CMOS fabrication may be considered as the price being paid for the improvements achieved in power consumption and noise margins.

In Fig. 5.16, note that the input voltage is connected to the gate terminals of both the nMOS and the pMOS transistors. Thus, both transistors are driven directly by the input signal, V_{in}. The substrate of the nMOS transistor is connected to the ground, while the substrate of the pMOS transistor is connected to the power supply voltage, V_{DD}, in order to reverse-bias the source and drain junctions. Since $V_{SB} = 0$ for both devices, there will be no substrate-bias effect for either device. It can be seen from the circuit diagram in Fig. 5.16 that

$$\begin{aligned} V_{GS,n} &= V_{in} \\ V_{DS,n} &= V_{out} \end{aligned} \tag{5.51}$$

and also,

$$\begin{aligned} V_{GS,p} &= -\left(V_{DD} - V_{in}\right) \\ V_{DS,p} &= -\left(V_{DD} - V_{out}\right) \end{aligned} \tag{5.52}$$

We will start our analysis by considering two simple cases. When the input voltage is smaller than the nMOS threshold voltage, i.e., when $V_{in} < V_{T0,n}$, the nMOS transistor is cut-off. At the same time, the pMOS transistor is on, operating in the linear region. Since the drain currents of both transistors are approximately equal to zero (except for small leakage currents), i.e.,

$$I_{D,n} = I_{D,p} = 0 \tag{5.53}$$

the drain-to-source voltage of the pMOS transistor is also equal to zero, and the output voltage V_{OH} is equal to the power supply voltage.

$$V_{out} = V_{OH} = V_{DD} \tag{5.54}$$

On the other hand, when the input voltage exceeds $(V_{DD} + V_{T0,p})$, the pMOS transistor is turned off. In this case, the nMOS transistor is operating in the linear region, but its drain-to-source voltage is equal to zero because condition (5.53) is satisfied. Consequently, the output voltage of the circuit is

$$V_{out} = V_{OL} = 0 \tag{5.55}$$

Next, we examine the operating modes of the nMOS and the pMOS transistors as functions of the input and output voltages. The nMOS transistor operates in *saturation* if $V_{in} > V_{T0,n}$ *and* if the following condition is satisfied.

$$V_{DS,n} \geq V_{GS,n} - V_{T0,n} \quad \Leftrightarrow \quad V_{out} \geq V_{in} - V_{T0,n} \tag{5.56}$$

The pMOS transistor operates in *saturation* if $V_{in} < (V_{DD} + V_{T0,p})$, *and* if :

$$V_{DS,p} \leq V_{GS,p} - V_{T0,p} \quad \Leftrightarrow \quad V_{out} \leq V_{in} - V_{T0,p} \tag{5.57}$$

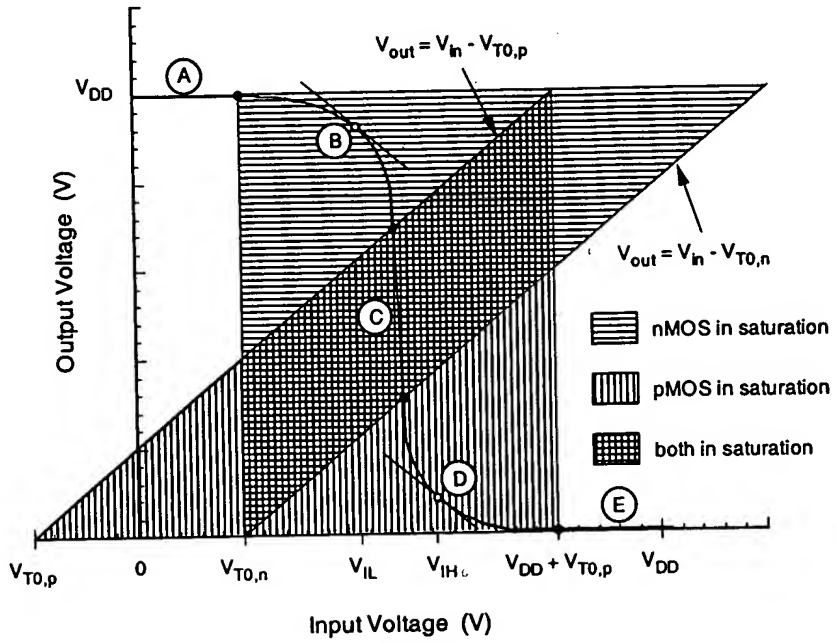


Figure 5.17. Operating regions of the nMOS and the pMOS transistors.

Both of these conditions for device saturation are illustrated graphically as shaded areas on the V_{out} - V_{in} plane in Fig. 5.17. A typical CMOS inverter voltage transfer characteristic is also superimposed for easy reference. Here, we identify five distinct regions, labeled A through E, each corresponding to a different set of operating conditions. The table below lists these regions and the corresponding critical input and output voltage levels.

Region	V_{in}	V_{out}	nMOS	pMOS
A	$< V_{T0,n}$	V_{OH}	cut-off	linear
B	V_{IL}	high $\approx V_{OH}$	saturation	linear
C	V_{th}	V_{th}	saturation	saturation
D	V_{IH}	low $\approx V_{OL}$	linear	saturation
E	$>(V_{DD} + V_{T0,p})$	V_{OL}	linear	cut-off

In Region A, where $V_{in} < V_{T0,n}$, the nMOS transistor is cut-off and the output voltage is equal to $V_{OH} = V_{DD}$. As the input voltage is increased beyond $V_{T0,n}$ (into Region B), the nMOS transistor starts conducting in saturation mode and the output voltage begins to decrease. Also note that the critical voltage V_{IL} which corresponds to $(dV_{out}/dV_{in}) = -1$ is located within Region B. As the output voltage further decreases, the pMOS transistor enters saturation at the boundary of Region C. It is seen from Fig. 5.17 that the inverter threshold voltage, where $V_{in} = V_{out}$, is located in Region C. When the output voltage V_{out} falls below $(V_{in} - V_{T0,n})$, the nMOS transistor starts to operate in linear mode. This corresponds to Region D in Fig. 5.17, where the critical voltage point V_{IH} with $(dV_{out}/dV_{in}) = -1$ is also located. Finally, in Region E, with the input voltage $V_{in} > (V_{DD} + V_{T0,p})$, the pMOS transistor is cut-off, and the output voltage is $V_{OL} = 0$.

In a simplistic analogy, the nMOS and the pMOS transistors can be seen as nearly ideal switches– controlled by the input voltage– that connect the output node to the power supply voltage or to the ground potential, depending on the input voltage level. The qualitative overview of circuit operation, illustrated in Fig. 5.17 and discussed above, also highlights the complementary nature of the CMOS inverter. The most significant feature of this circuit is that the current drawn from the power supply in both of these steady-state operating points, i.e., in Region A and in Region E, is nearly equal to zero. The only current that flows in either case is the very small leakage current of the reverse-biased source and drain junctions. The CMOS inverter can *drive* any load, such as interconnect capacitance or fanout logic gates which are connected to its output node, either by supplying current to the load, or by sinking current from the load.

The steady-state input-output voltage characteristics of the CMOS inverter can be better visualized by considering the interaction of individual nMOS and pMOS transistor characteristics in the current-voltage space. We already know that the drain current $I_{D,n}$ of the nMOS transistor is a function of the voltages $V_{GS,n}$ and $V_{DS,n}$. Hence, the nMOS drain current is also a function of the inverter input and output voltages V_{in} and V_{out}, according to (5.51).

$$I_{D,n} = f(V_{in}, V_{out})$$

This two-variable function, which is essentially described by the current equations (3.54) through (3.56), can be represented as a *surface* in the three-dimensional current-voltage space. Figure 5.18 shows this $I_{D,n}(V_{in}, V_{out})$ surface for the nMOS transistor.

Similarly, the drain current $I_{D,p}$ of the pMOS transistor is also a function of the inverter input and output voltages V_{in} and V_{out}, according to (5.52).

$$I_{D,p} = f(V_{in}, V_{out})$$

This two-variable function, described by the current equations (3.57) through (3.59), can be represented as another surface in the three-dimensional current-voltage space. Figure 5.19 shows the corresponding $I_{D,p}(V_{in}, V_{out})$ surface for the pMOS transistor.

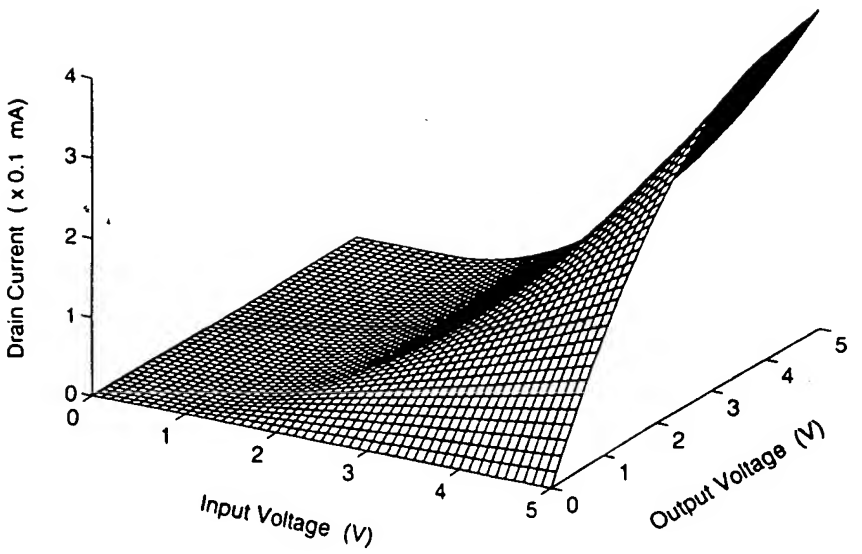


Figure 5.18. Current-voltage surface representing the nMOS transistor characteristics.

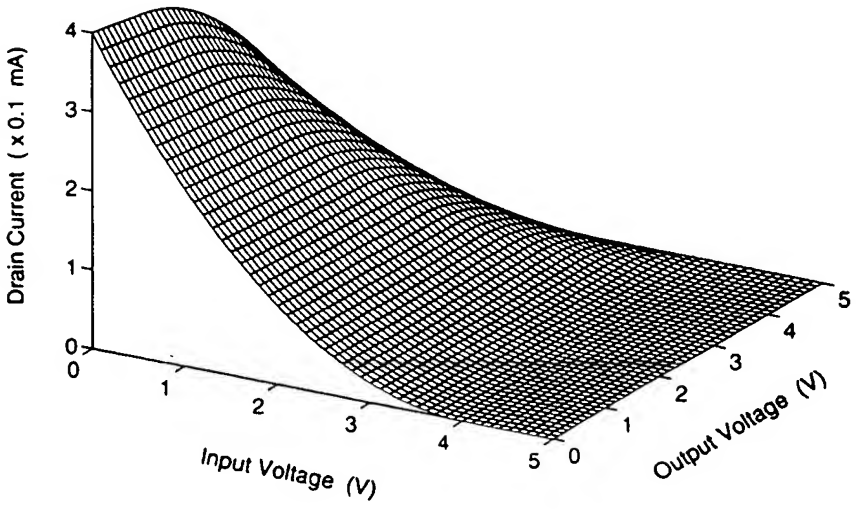


Figure 5.19. Current-voltage surface representing the pMOS transistor characteristics.

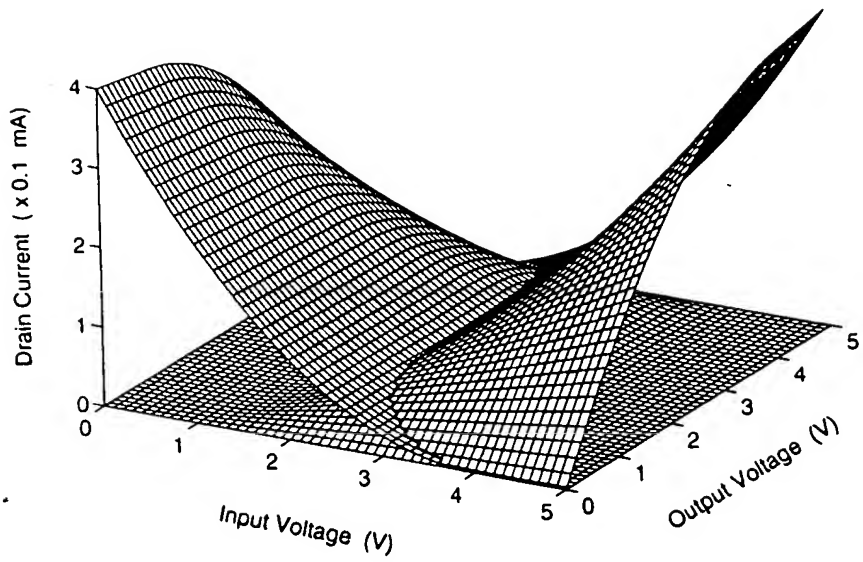


Figure 5.20. Intersection of the current-voltage surfaces shown in Figures 5.18 and 5.19.

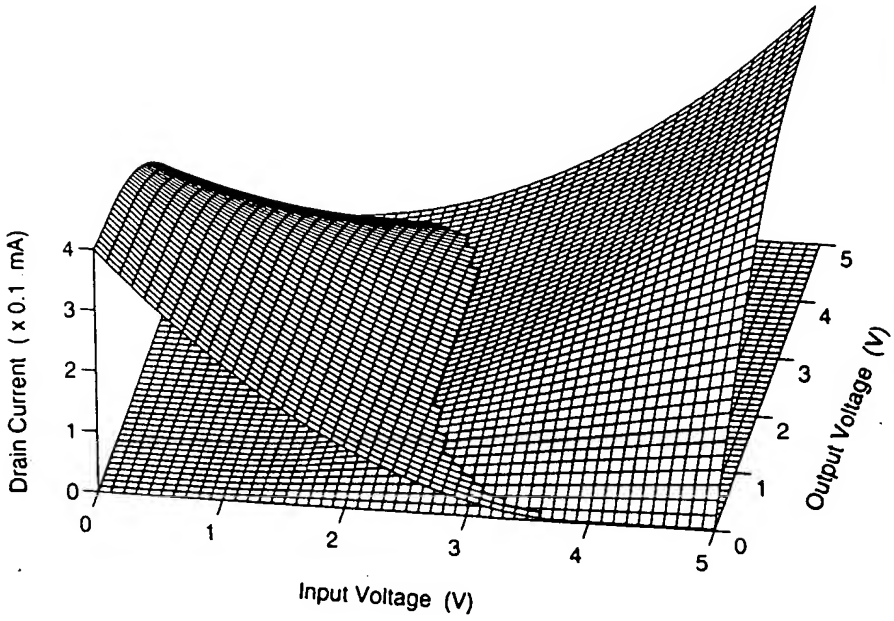


Figure 5.21. The intersecting current-voltage surfaces shown from a different viewing angle. Notice that projection of the intersection curve on the voltage plane gives the VTC.

Remember that in a CMOS inverter operating in steady-state, the drain current of the nMOS transistor is always equal to the drain current of the pMOS transistor, according to KCL.

$$I_{D,n} = I_{D,p}$$

Thus, the *intersection* of the two current-voltage surfaces shown in Figs. 5.18 and 5.19 will give the operating curve of the CMOS inverter circuit in the three-dimensional current-voltage space. The intersection of the two characteristic surfaces is shown in Fig. 5.20. The intersecting surfaces are shown from a different viewing angle in Fig. 5.21, with the intersection curve highlighted in bold.

It is clear that the vertical projection of the intersection curve on the V_{in} - V_{out} plane produces the typical CMOS inverter voltage transfer characteristic already shown in Fig. 5.17. Similarly, the horizontal projection of the intersection curve on the I_D - V_{in} plane gives the steady-state current drawn by the inverter from the power supply voltage as a function of the input voltage. In the following, we will present an in-depth analysis of the CMOS inverter static characteristics, by calculating the critical voltage points on the VTC. It has already been established that $V_{OH} = V_{DD}$ and $V_{OL} = 0$ for this inverter; thus, we will devote our attention to V_{IL}, V_{IH} and the inverter switching threshold, V_{th}.

Calculation of V_{IL}

By definition, the slope of the VTC is equal to (−1), i.e., $dV_{out} / dV_{in} = -1$ when the input voltage is $V_{in} = V_{IL}$. Note that in this case, the nMOS transistor operates in saturation while the pMOS transistor operates in the linear region. From $I_{D,n} = I_{D,p}$, we obtain the following current equation:

$$\frac{k_n}{2}\cdot\left(V_{GS,n} - V_{T0,n}\right)^2 = \frac{k_p}{2}\cdot\left[2\cdot\left(V_{GS,p} - V_{T0,p}\right)\cdot V_{DS,p} - {V_{DS,p}}^2\right] \tag{5.58}$$

Using equations (5.51) and (5.52), this expression can be rewritten as

$$\frac{k_n}{2}\cdot\left(V_{in} - V_{T0,n}\right)^2 = \frac{k_p}{2}\cdot\left[2\cdot\left(V_{in} - V_{DD} - V_{T0,p}\right)\cdot\left(V_{out} - V_{DD}\right) - \left(V_{out} - V_{DD}\right)^2\right] \tag{5.59}$$

To satisfy the derivative condition at V_{IL}, we differentiate both sides of (5.59) with respect to V_{in}.

$$k_n\cdot\left(V_{in} - V_{T0,n}\right) = k_p\cdot\left[\left(V_{in} - V_{DD} - V_{T0,p}\right)\cdot\left(\frac{dV_{out}}{dV_{in}}\right) + \left(V_{out} - V_{DD}\right) - \left(V_{out} - V_{DD}\right)\cdot\left(\frac{dV_{out}}{dV_{in}}\right)\right] \tag{5.60}$$

Substituting $V_{in} = V_{IL}$ and $(dV_{out}/dV_{in}) = -1$ in (5.60), we obtain

$$k_n \cdot \left(V_{IL} - V_{T0,n}\right) = k_p \cdot \left(2\,V_{out} - V_{IL} + V_{T0,p} - V_{DD}\right) \tag{5.61}$$

The critical voltage V_{IL} can now be found as a function of the output voltage V_{out}, as follows:

$$V_{IL} = \frac{2\,V_{out} + V_{T0,p} - V_{DD} + k_R\,V_{T0,n}}{1 + k_R} \tag{5.62}$$

where k_R is defined as

$$k_R = \frac{k_n}{k_p}$$

This equation must be solved together with the KCL equation (5.59) to obtain the numerical value of V_{IL} and the corresponding output voltage, V_{out}. Note that the solution is fairly straightforward and does not require numerical iterations as in the previous cases, since none of the transistors is subject to substrate-bias effects.

Calculation of V_{IH}

When the input voltage is equal to V_{IH}, the nMOS transistor operates in the linear region, and the pMOS transistor operates in saturation. Applying KCL to the output node, we obtain

$$\frac{k_n}{2} \cdot \left[2 \cdot \left(V_{GS,n} - V_{T0,n}\right) \cdot V_{DS,n} - {V_{DS,n}}^2\right] = \frac{k_p}{2} \cdot \left(V_{GS,p} - V_{T0,p}\right)^2 \tag{5.63}$$

Using equations (5.51) and (5.52), this expression can be rewritten as

$$\frac{k_n}{2} \cdot \left[2 \cdot \left(V_{in} - V_{T0,n}\right) \cdot V_{out} - {V_{out}}^2\right] = \frac{k_p}{2} \cdot \left(V_{in} - V_{DD} - V_{T0,p}\right)^2 \tag{5.64}$$

Now, differentiate both sides of (5.64) with respect to V_{in}.

$$k_n \cdot \left[\left(V_{in} - V_{T0,n}\right) \cdot \left(\frac{dV_{out}}{dV_{in}}\right) + V_{out} - V_{out} \cdot \left(\frac{dV_{out}}{dV_{in}}\right)\right] = k_p \cdot \left(V_{in} - V_{DD} - V_{T0,p}\right) \tag{5.65}$$

Substituting $V_{in} = V_{IH}$ and $(dV_{out}/dV_{in}) = -1$ in (5.65), we obtain

$$k_n \cdot \left(-V_{IH} + V_{T0,n} + 2V_{out}\right) = k_p \cdot \left(V_{IH} - V_{DD} - V_{T0,p}\right) \tag{5.66}$$

The critical voltage V_{IH} can now be found as a function of V_{out} as follows:

$$V_{IH} = \frac{V_{DD} + V_{T0,p} + k_R \cdot \left(2V_{out} + V_{T0,n}\right)}{1 + k_R} \tag{5.67}$$

Again, this equation must be solved simultaneously with the KCL equation (5.64) to obtain the numerical values of V_{IH} and V_{out}.

Calculation of V_{th}

The inverter threshold voltage is defined as $V_{th} = V_{in} = V_{out}$. Since the CMOS inverter exhibits large noise margins and a very sharp VTC transition, the inverter threshold voltage emerges as an important parameter characterizing the DC performance of the inverter. For $V_{in} = V_{qut}$, both transistors are expected to be in saturation mode; hence, we can write the following KCL equation.

$$\frac{k_n}{2} \cdot \left(V_{GS,n} - V_{T0,n}\right)^2 = \frac{k_p}{2} \cdot \left(V_{GS,p} - V_{T0,p}\right)^2 \tag{5.68}$$

Replacing $V_{GS,n}$ and $V_{GS,p}$ in (5.68) according to (5.51) and (5.52), we obtain

$$\frac{k_n}{2} \cdot \left(V_{in} - V_{T0,n}\right)^2 = \frac{k_p}{2} \cdot \left(V_{in} - V_{DD} - V_{T0,p}\right)^2 \tag{5.69}$$

The correct solution for V_{in} for this equation is

$$V_{in} \cdot \left(1 + \sqrt{\frac{k_p}{k_n}}\right) = V_{T0,n} + \sqrt{\frac{k_p}{k_n}} \cdot \left(V_{DD} + V_{T0,p}\right) \tag{5.70}$$

Finally, the inverter threshold (switching threshold) voltage V_{th} is found as

$$V_{th} = \frac{V_{T0,n} + \sqrt{\frac{1}{k_R}} \cdot \left(V_{DD} + V_{T0,p}\right)}{\left(1 + \sqrt{\frac{1}{k_R}}\right)} \tag{5.71}$$

Note that the inverter threshold voltage is defined as $V_{th} = V_{in} = V_{out}$. When the input voltage is equal to V_{th}, however, we find that the output voltage can actually attain any value between $(V_{th} - V_{T0,n})$ and $(V_{th} - V_{T0,p})$, without violating the voltage conditions used in this analysis. This is due to the fact that the VTC segment corresponding to Region C in Fig. 5.17 becomes completely vertical if the channel-length modulation effect is neglected, i.e., if $\lambda = 0$. In more realistic cases with $\lambda > 0$, the VTC segment in Region C exhibits a finite, but very large, slope. Figure 5.22 shows the variation of the inversion (switching) threshold voltage V_{th} as a function of the transconductance ratio k_R, and for fixed values of V_{DD}, $V_{T0,n}$ and $V_{T0,p}$.

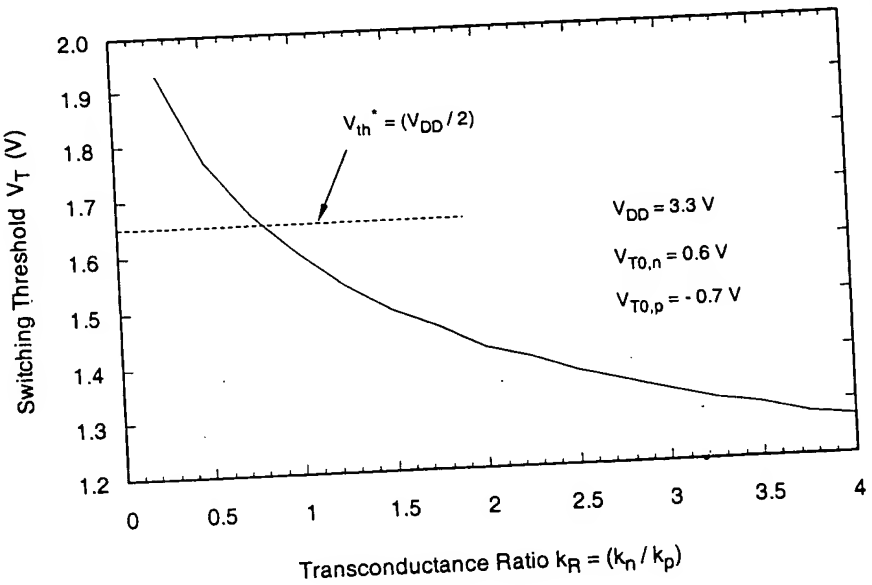


Figure 5.22. Variation of the inversion threshold voltage as a function of k_R.

It has already been established that the CMOS inverter does not draw any significant current from the power source, except for small leakage and subthreshold currents, when the input voltage is either smaller than $V_{T0,n}$ or larger than $(V_{DD}+V_{T0,p})$. The nMOS and the pMOS transistors conduct a nonzero current, on the other hand, during low-to-high and high-to-low transitions, i.e., in Regions B, C, and D. It can be shown that the current being drawn from the power source during transition reaches its peak value when $V_{in} = V_{th}$. In other words, the maximum current is drawn when both transistors are operating in saturation mode. Figure 5.23 shows the voltage transfer characteristic of a typical CMOS inverter circuit and the power supply current, as a function of the input voltage.

Design of CMOS Inverters

The inverter threshold voltage V_{th} was identified as one of the most important parameters that characterize the steady-state input-output behavior of the CMOS inverter circuit. The

CMOS inverter can, by virtue of its complementary push-pull operating mode, provide a full output voltage swing between 0 and V_{DD}, and therefore, the noise margins are relatively wide. Thus, the problem of designing a CMOS inverter can be reduced to setting the inverter threshold to a desired voltage value.

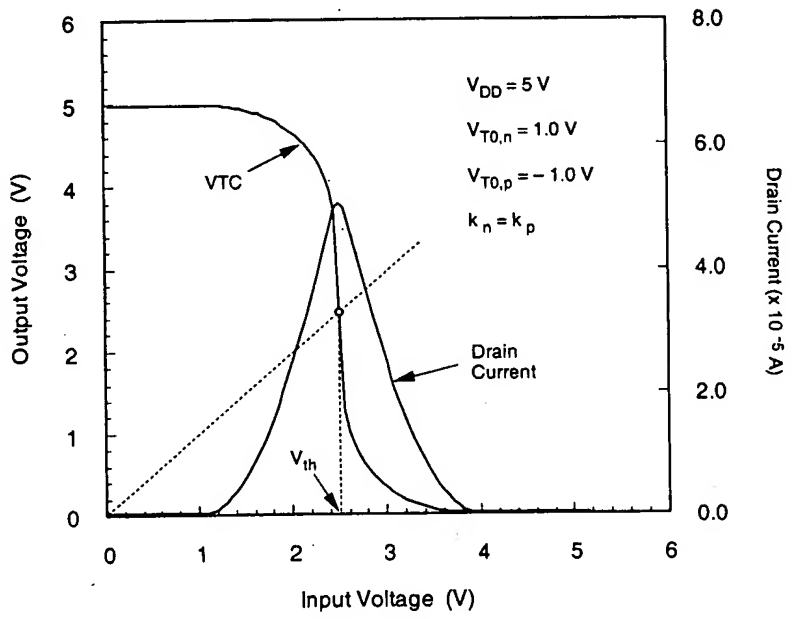


Figure 5.23. Typical VTC and the power supply current of a CMOS inverter circuit.

Given the power supply voltage V_{DD}, the nMOS and the pMOS transistor threshold voltages, and the desired inverter threshold voltage V_{th}, the corresponding ratio k_R can be found as follows. Reorganizing (5.71) yields

$$\sqrt{\frac{1}{k_R}} = \frac{V_{th} - V_{T0,n}}{V_{DD} + V_{T0,p} - V_{th}} \tag{5.72}$$

Now solve for k_R that is required to achieve the given V_{th}.

$$k_R = \frac{k_n}{k_p} = \left(\frac{V_{DD} + V_{T0,p} - V_{th}}{V_{th} - V_{T0,n}} \right)^2 \tag{5.73}$$

Recall that the switching threshold voltage of an *ideal* inverter is defined as

$$V_{th,ideal} = \frac{1}{2} \cdot V_{DD} \tag{5.74}$$

Substituting (5.74) in (5.73) gives

$$\left(\frac{k_n}{k_p}\right)_{ideal} = \left(\frac{0.5\,V_{DD} + V_{T0,p}}{0.5\,V_{DD} - V_{T0,n}}\right)^2 \tag{5.75}$$

for a near-ideal CMOS VTC that satisfies the condition (5.74). Since the operations of the nMOS and the pMOS transistors of the CMOS inverter are fully complementary, we can achieve completely symmetric input-output characteristics by setting the threshold voltages as $V_{T0} = V_{T0,n} = |V_{T0,p}|$. This reduces (5.75) to:

$$\left(\frac{k_n}{k_p}\right)_{\substack{symmetric \\ inverter}} = 1 \tag{5.76}$$

Note that the ratio k_R is defined as

$$\frac{k_n}{k_p} = \frac{\mu_n C_{ox} \cdot \left(\frac{W}{L}\right)_n}{\mu_p C_{ox} \cdot \left(\frac{W}{L}\right)_p} = \frac{\mu_n \cdot \left(\frac{W}{L}\right)_n}{\mu_p \cdot \left(\frac{W}{L}\right)_p} \tag{5.77}$$

assuming that the gate oxide thickness t_{ox}, and hence, the gate oxide capacitance C_{ox} have the same value for both nMOS and pMOS transistors. The unity-ratio condition (5.76) for the ideal symmetric inverter requires that

$$\frac{\left(\frac{W}{L}\right)_n}{\left(\frac{W}{L}\right)_p} = \frac{\mu_p}{\mu_n} \approx \frac{230\ \text{cm}^2/\text{V}\cdot\text{s}}{580\ \text{cm}^2/\text{V}\cdot\text{s}} \tag{5.78}$$

Hence,

$$\left(\frac{W}{L}\right)_p \approx 2.5\left(\frac{W}{L}\right)_n \tag{5.79}$$

It should be noted that the numerical values used in (5.78) for electron and hole mobilities are *typical* values, and that exact μ_n and μ_p values will vary with surface doping concentration of the substrate and the tub. The VTCs of three CMOS inverter circuits with different k_R ratios are shown in Fig. 5.24. It can be seen clearly that the inverter threshold voltage V_{th} shifts to lower values with increasing k_R ratio.

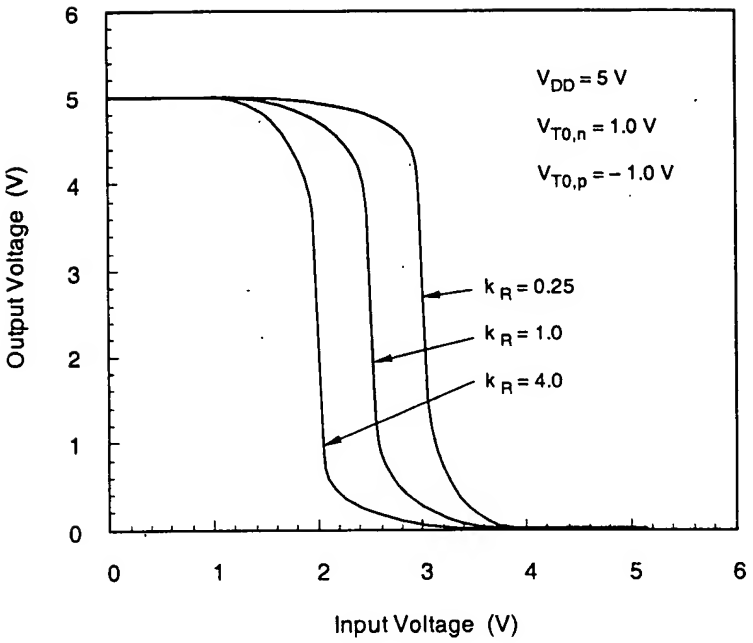


Figure 5.24. Voltage transfer characteristics of three CMOS inverters, with different nMOS-to-pMOS ratios.

For a symmetric CMOS inverter with $V_{T0,n} = |V_{T0,p}|$ and $k_R = 1$, the critical voltage V_{IL} can be found, using (5.62), as follows:

$$V_{IL} = \frac{1}{8} \cdot \left(3\,V_{DD} + 2\,V_{T0,n}\right) \tag{5.80}$$

Also, the critical voltage V_{IH} is found as

$$V_{IH} = \frac{1}{8} \cdot \left(5\,V_{DD} - 2\,V_{T0,n}\right) \tag{5.81}$$

Note that the sum of V_{IL} and V_{IH} is always equal to V_{DD} in a symmetric inverter.

$$V_{IL} + V_{IH} = V_{DD} \tag{5.82}$$

The noise margins NM_L and NM_H for this symmetric CMOS inverter are now calculated using (5.3) and (5.4).

$$NM_L = V_{IL} - V_{OL} = V_{IL}$$
$$NM_H = V_{OH} - V_{IH} = V_{DD} - V_{IH} \tag{5.83}$$

which are equal to each other, and also to V_{IL}.

$$NM_L = NM_H = V_{IL} \tag{5.84}$$

Example 5.4

Consider a CMOS inverter circuit with the following parameters :

$V_{DD} = 3.3$ V
$V_{T0,n} = 0.6$ V
$V_{T0,p} = -0.7$ V
$k_n = 200\ \mu A/V^2$
$k_p = 80\ \mu A/V^2$

Calculate the noise margins of the circuit. Notice that the CMOS inverter being considered here has $k_R = 2.5$ and $V_{T0,n} \neq |V_{T0,p}|$; hence, it is not a symmetric inverter.

First, the output low voltage V_{OL} and the output high voltage V_{OH} are found, using (5.54) and (5.55), as $V_{OL} = 0$ and $V_{OH} = 5$ V. To calculate V_{IL} in terms of the output voltage, we use (5.62).

$$V_{IL} = \frac{2V_{out} + V_{T0,p} - V_{DD} + k_R V_{T0,n}}{1 + k_R}$$
$$= \frac{2V_{out} - 0.7 - 3.3 + 1.5}{1 + 2.5} = 0.57 V_{out} - 0.71$$

Now substitute this expression into the KCL equation (5.59).

$$2.5(0.57 V_{out} - 0.71 - 0.6)^2 = 2(0.57 V_{out} - 0.71 - 3.3 + 0.7)(V_{out} - 3.3) - (V_{out} - 3.3)^2$$

This expression yields a second-order polynomial in V_{out}, as follows:

$$0.66 V_{out}^2 + 0.05 V_{out} - 6.65 = 0$$

Only one root of this quadratic equation corresponds to a physically correct solution for V_{out} (i.e., $V_{out} > 0$).

$$V_{out} = 3.14 \text{ V}$$

From this value, we can calculate the critical voltage V_{IL} as:

$$V_{IL} = 0.57 \cdot 3.14 - 0.71 = \underline{\underline{1.08 \text{ V}}}$$

To calculate V_{IH} in terms of the output voltage, use (5.67):

$$V_{IH} = \frac{V_{DD} + V_{T0,p} + k_R \cdot \left(2 V_{out} + V_{T0,n}\right)}{1 + k_R}$$

$$= \frac{3.3 - 0.7 + 2.5\left(2 V_{out} + 0.6\right)}{1 + 2.5} = 1.43 V_{out} + 1.17$$

Next, substitute this expression into the KCL equation (5.64) to obtain a second-order polynomial in V_{out}.

$$2.5\left[2\left(1.43 V_{out} + 1.17 - 0.6\right) V_{out} - {V_{out}}^2\right] = \left(1.43 V_{out} - 1.43\right)^2$$

$$2.61 {V_{out}}^2 + 6.94 V_{out} - 2.04 = 0$$

Again, only one root of this quadratic equation corresponds to the physically correct solution for V_{out} at this operating point, i.e., when $V_{in} = V_{IH}$.

$$V_{out} = 0.27 \text{ V}$$

From this value, we can calculate the critical voltage V_{IH} as:

$$V_{IH} = 1.43 \cdot 0.37 + 1.17 = \underline{\underline{1.55 \text{ V}}}$$

Finally, we find the noise margins for low voltage levels and for high voltage levels using (5.3) and (5.4).

$$NM_L = V_{IL} - V_{OL} = 1.08 \text{ V}$$
$$NM_H = V_{OH} - V_{IH} = 1.75 \text{ V}$$

Supply Voltage Scaling in CMOS Inverters

In the following, we will briefly examine the effects of supply voltage scaling, i.e., reduction of V_{DD}, upon the static voltage transfer characteristics of CMOS inverters. The

overall power dissipation of any digital circuit is a strong function of the supply voltage V_{DD}. With the growing trend for reducing the power dissipation in large-scale integrated systems and especially in portable applications, reduction (or scaling) of the power supply voltage emerges as one of the most widely practiced measures for low-power design. While such reduction is usually very effective, several important issues must also be addessed so that the system performance is not sacrificed. In this context, it is quite relevant to explore the influence of supply voltage scaling upon the VTC of simple CMOS inverter circuits.

The expressions we have developed in this section for V_{IL}, V_{IH}, and V_{th} indeed show that the static characteristics of the CMOS inverter allow significant variation of the supply voltage without affecting the functionality of the basic inverter. Neglecting second-order effects such as subthreshold conduction, it can be seen that the CMOS inverter will continue to operate correctly with a supply voltage which is as low as the following limit value.

$$V_{DD}{}^{min} = V_{T0,n} + |V_{T0,p}| \tag{5.85}$$

This means that correct inverter operation will be sustained if at least one of the transistors remains in conduction, for any given input voltage. Figure 5.25 shows the voltage transfer characteristics of a CMOS inverter, obtained with different supply voltage levels. The exact shape of the VTC near the limit value is essentially determined by subthreshold conduction properties of the nMOS and pMOS transistors, yet it is clear that the circuit operates as an inverter over a large range of supply voltages levels.

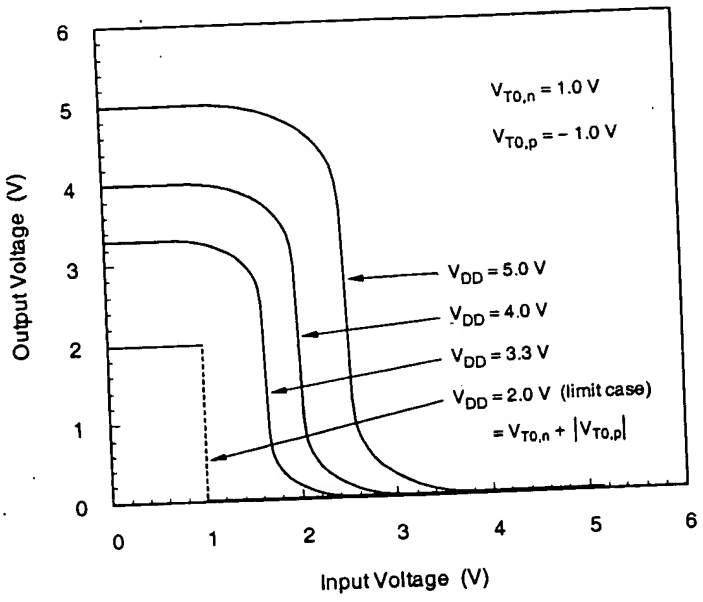


Figure 5.25. Voltage transfer characteristics of a CMOS inverter, obtained with different power supply voltage levels.

It is interesting to note that the CMOS inverter will continue to operate, albeit somewhat differently, even beyond the limit described in (5.85). If the power supply voltage is reduced *below* the sum of the two threshold voltages, the VTC will contain a region in which none of the transistors is conductiong. The output voltage level within this region is then determined by the *previous state* of the output, since the previous output level is always preserved as stored charge at the output node. Thus, the VTC exhibits a *hysteresis* behavior for very low supply voltage levels, which is illustrated in Fig. 5.26.

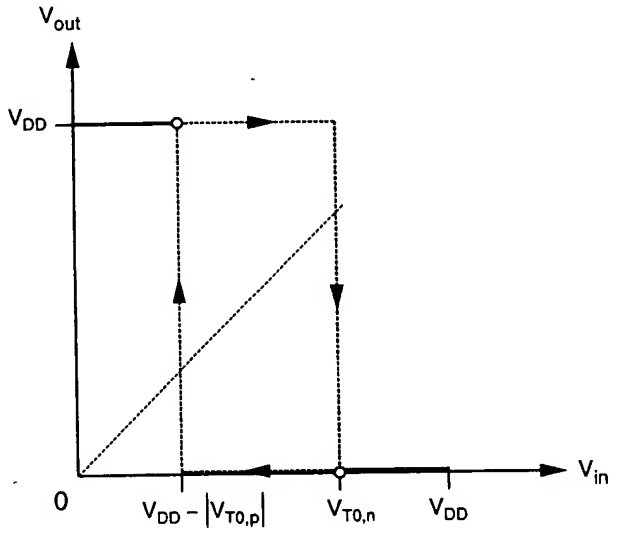


Figure 5.26. Voltage transfer characteristic of a CMOS inverter, operated with a supply voltage which is lower than the limit given in (5.85).

Power and Area Considerations

Since the CMOS inverter does not draw any significant current from the power source in both of its steady-state operating points ($V_{out} = V_{OH}$ and $V_{out} = V_{OL}$), the DC power dissipation of this circuit is almost negligible. The drain current that flows through the nMOS and the pMOS transistors in both cases is essentially limited to the reverse leakage current of the source and drain pn-junctions, and in short-channel MOSFETs, the relatively small subthreshold current. This unique property of the CMOS inverter was already identified as one of the most important advantages of this configuration. In many applications requiring a low overall power consumption, CMOS is preferred over other circuit alternatives for this reason. It must be noted, however, that the CMOS inverter does conduct a significant amount of current during a *switching event*, i.e., when the output voltage changes from a low to high state, or from a high to low state. The detailed calculation of this *dynamic power dissipation* will be examined in Chapters 6 and 11.

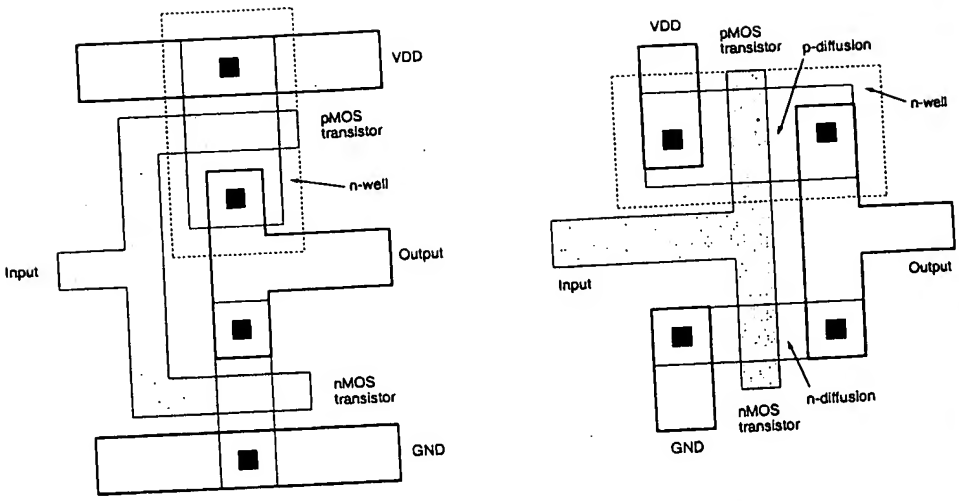


Figure 5.27. Two sample layouts of CMOS inverter circuits (for p-type substrate).

Figure 5.27 shows two layout examples for the simple CMOS inverter circuit. In both cases, it is assumed that the circuit is being built on a *p*-type wafer, which also provides the substrate for the nMOS transistor. The pMOS transistor, on the other hand, must be placed in an n-well (dotted lines), which becomes the substrate for this device. Also note that in Fig. 5.27 the channel width of the pMOS transistor is larger than that of the nMOS transistor. This is typical for symmetric inverter configurations, in which the k_R ratio is set approximately equal to unity.

Compared to other inverter layouts examined in previous sections, the CMOS inverters shown in Fig. 5.27 do not occupy significantly more area. The added complexity of the fabrication process (creating n-well diffusion, separate p-type and n-type source and drain diffusions, etc.) appears to be the only drawback for the inverter example. Because of the complementary nature of this circuit configuration, however, CMOS random logic circuits require significantly more transistors for the same function than their nMOS counterparts. Consequently, CMOS logic circuits tend to occupy more area than comparable nMOS logic circuits, which apparently affects the integration density of pure-CMOS logic. The actual integration density of nMOS logic, on the other hand, is limited by power dissipation and heat generation problems.

References

1. N.H.E. Weste and K. Eshraghian, *Principles of CMOS VLSI Design–A Systems Perspective*, second edition, Reading, MA: Addison-Wesley, 1993.

2. J.P. Uyemura, *Fundamentals of MOS Digital Integrated Circuits*, Reading, MA: Addison-Wesley, 1988.

It can be seen that the propagation delay of the lumped RC network is about 2.5 ns (regardless of signal flow direction), while the propagation delay of the 10-segment RC ladder network is about 0.7 ns (for signal flow from B to A). In this case, the difference between the delay times is significantly more pronounced due to the nonuniform distribution of parasitics. As in the previous example, we also construct the simple T-model of this nonuniform interconnect line, consisting of one lumped capacitor (356 fF) and two lumped resistors (2.5 kΩ and 7.5 kΩ). The T-model again yields a significantly more accurate transient response, and it correctly represents the directional dependence of propagation delay times which is due to the nonuniform geometry of the line.

6.7. Switching Power Dissipation of CMOS Inverters

It was shown in Chapter 5 that the static power dissipation of the CMOS inverter is quite negligible. During switching events where the output load capacitance is alternatingly charged up and charged down, on the other hand, the CMOS inverter inevitably dissipates power. In the following section, we will derive the expressions for the dynamic power consumption of the CMOS inverter.

Consider the simple CMOS inverter circuit shown in Fig. 6.27. We will assume that the input voltage is an ideal step waveform with negligible rise and fall times. Typical input and output voltage waveforms and the expected load capacitor current waveform are shown in Fig. 6.28. When the input voltage switches from low to high, the pMOS transistor in the circuit is turned off, and the nMOS transistor starts conducting. During this phase, the output load capacitance C_{load} is being discharged through the nMOS transistor. Thus, the capacitor current equals the instantaneous drain current of the nMOS transistor. When the input voltage switches from high to low, the nMOS transistor in the circuit is turned off, and the pMOS transistor starts conducting. During this phase, the output load capacitance C_{load} is being charged up through the pMOS transistor; therefore, the capacitor current equals the instantaneous drain current of the pMOS transistor.

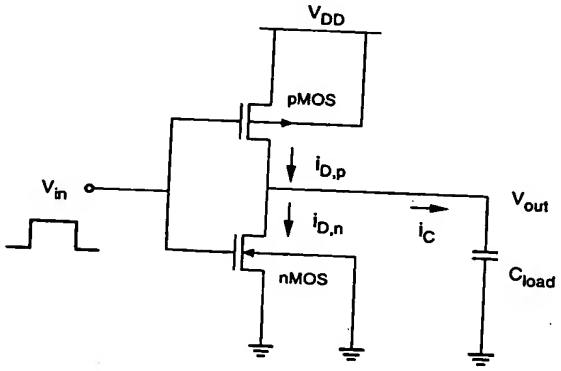


Figure 6.27. CMOS inverter used in the dynamic power-dissipation analysis.

Assuming periodic input and output waveforms, the average power dissipated by any device over one period can be found as follows:

$$P_{avg} = \frac{1}{T}\int_0^T v(t)\cdot i(t)\,dt \tag{6.66}$$

Since during switching, the nMOS transistor and the pMOS transistor in a CMOS inverter conduct current for one-half period each, the average power dissipation of the CMOS inverter can be calculated as the power required to charge up and charge down the output load capacitance.

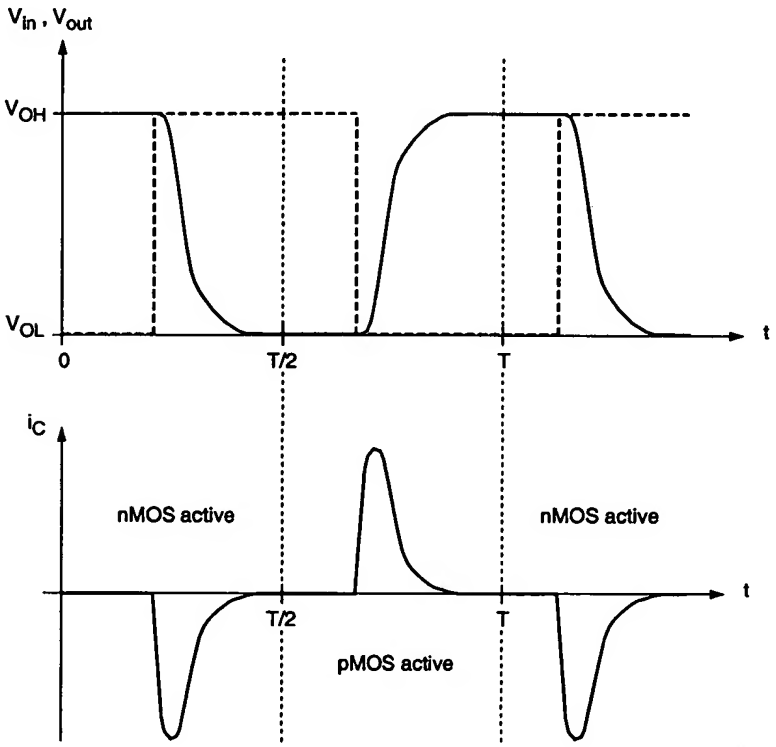


Figure 6.28. Typical input and output voltage waveforms and the capacitor current waveform during switching of the CMOS inverter.

$$P_{avg} = \frac{1}{T}\left[\int_0^{T/2} V_{out}\left(-C_{load}\frac{dV_{out}}{dt}\right)dt + \int_{T/2}^{T}\left(V_{DD}-V_{out}\right)\left(C_{load}\frac{dV_{out}}{dt}\right)dt\right] \tag{6.67}$$

Evaluating the integrals in (6.67), we obtain

$$P_{avg} = \frac{1}{T}\left[\left(-C_{load}\frac{V_{out}^2}{2}\right)\Big|_0^{T/2} + \left(V_{DD}\cdot V_{out}\cdot C_{load} - \frac{1}{2}C_{load}V_{out}^2\right)\Big|_{T/2}^{T}\right] \tag{6.68}$$

$$P_{avg} = \frac{1}{T}C_{load}V_{DD}^2 \tag{6.69}$$

Noting that $f = 1/T$, this expression can also be written as:

$$P_{avg} = C_{load}\cdot V_{DD}^2 \cdot f \tag{6.70}$$

It is clear that the average power dissipation of the CMOS inverter is proportional to the switching frequency f. Therefore, the low-power advantage of CMOS circuits becomes less prominent in high-speed operation, where the switching frequency is high. Also note that the average power dissipation is independent of all transistor characteristics and transistor sizes. Consequently, the switching delay times have no relevance to the amount of power consumption during the switching events. The reason for this is that the switching power is solely dissipated for charging and discharging the output capacitance from V_{OL} to V_{OH}, and vice versa.

For this reason, the switching power expression derived for the CMOS inverter also applies to all general CMOS circuits, as shown in Fig. 6.29. A general CMOS logic circuit

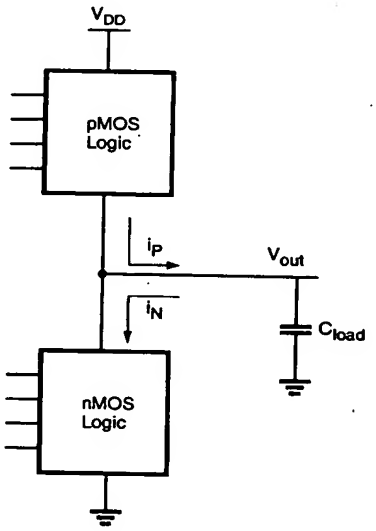


Figure 6.29. Generalized CMOS logic circuit.

consists of an nMOS logic block between the output node and the ground, and a pMOS logic block between the output and V_{DD}. As in the simple CMOS inverter case, either the pMOS block *or* the nMOS block can conduct depending on the input voltage combination, but not both at the same time. Therefore, switching power is again dissipated solely for charging and discharging the output capacitance.

To summarize, if the total parasitic capacitance in the circuit can be lumped at the output node with reasonable accuracy, if the output voltage swing is between 0 and V_{DD}, and if the input voltage waveforms are assumed to be ideal step inputs, the average switching power expression (6.70) will hold for any CMOS logic circuit.

Note that under realistic conditions, when the input voltage waveform deviates from ideal step input and has nonzero rise and fall times, for example, both the nMOS and the pMOS transistor will simultaneously conduct a certain amount of current during the switching event. This is called the short-circuit current, since in this case, the two transistors temporarily form a conducting path between the V_{DD} and the ground. The additional power dissipation, which is due to the short-circuit current, cannot be predicted by the power-dissipation formula (6.70) derived above, since the short-circuit current is not being utilized to charge or discharge the output load capacitor. We must be aware that this additional power-dissipation term can be quite significant under some nonideal conditions. If the load capacitance is increased, on the other hand, the short-circuit dissipation term usually becomes negligible in comparison to the power dissipation which is due to the charging/discharging of capacitances.

Power Meter Simulation

In the following, we present a simple circuit simulation approach which can be used to estimate the average power dissipation of arbitrary circuits (including the effects of short circuit and leakage currents), under realistic operating conditions. According to (6.66), the average power dissipation of any device or circuit which is driven by a periodic input waveform can be found by integrating the product of its instantaneous terminal voltage and its instantaneous terminal current over one period. If we have to determine the amount of P_{avg} drawn from the power supply over one period, the problem is reduced to finding only the time-average of the power supply current, since the power supply voltage is a constant.

Using a simple simulation model called the *power meter*, we can estimate the average power dissipation of an arbitrary device or circuit driven by a periodic input, with transient circuit simulation. Consider the circuit structure shown in Fig. 6.30, in which a zero-volt independent voltage source is connected in series with the power supply voltage source V_{DD} of the device or circuit in question. Consequently, the instantaneous power supply current $i_{DD}(t)$ which is being drawn by the circuit will also pass through the zero-volt voltage source, $i_s(t) = i_{DD}(t)$.

The power meter circuit consists of three elements: a linear current-controlled current source, a capacitor, and a resistor, all connected in parallel. The current equation for the common node of the power meter circuit can be written as follows:

$$C_y \frac{dV_y}{dt} = \beta i_S - \frac{V_y}{R_y} \tag{6.71}$$

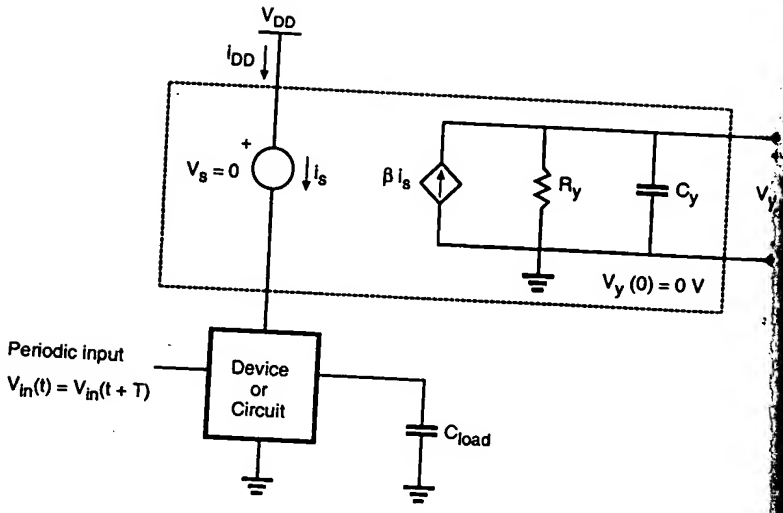


Figure 6.30. The power meter circuit used for the simulation of average dynamic dissipation of an arbitrary device or circuit.

The initial condition of the node voltage V_y is set as $V_y(0) = 0$ V. Then, the time-d solution of $V_y(t)$ can be found by integrating (6.71).

$$V_y(t) = \frac{\beta}{C_y} \int_0^t e^{-\frac{1}{R_y C_y}(t-\tau)} i_{DD}(\tau)\, d\tau$$

Assuming $R_y C_y >> T$, the voltage value $V_y(T)$ at the end of one period can be approxi as follows.

$$V_y(T) \approx \frac{\beta}{C_y} \int_0^T i_{DD}(\tau)\, d\tau$$

If the constant coefficient value of the current-controlled current source is set to

$$\beta = V_{DD} \frac{C_y}{T}$$

the voltage value $V_y(T)$ at the end of one period will be found by transient simula

$$V_y(T) = V_{DD} \cdot \frac{1}{T} \int_0^T i_{DD}(\tau)\, d\tau$$

Note that the right-hand side of (6.75) corresponds to the average power drawn from the power supply source over one period. Thus, the value of the node voltage V_y at $t = T$ gives the average power dissipation.

The power meter circuit shown in Fig. 6.30 can be easily simulated using a conventional circuit simulation program such as SPICE, and it enables us to accurately estimate the average power dissipation of any circuit with arbitrary complexity. Also note that the power meter circuit inherently takes into account the additional power dissipation due to the short-circuit currents, which may arise because of nonideal input conditions. In the following example, we present a sample SPICE simulation of the power meter for estimating the dynamic power dissipation of a CMOS inverter circuit.

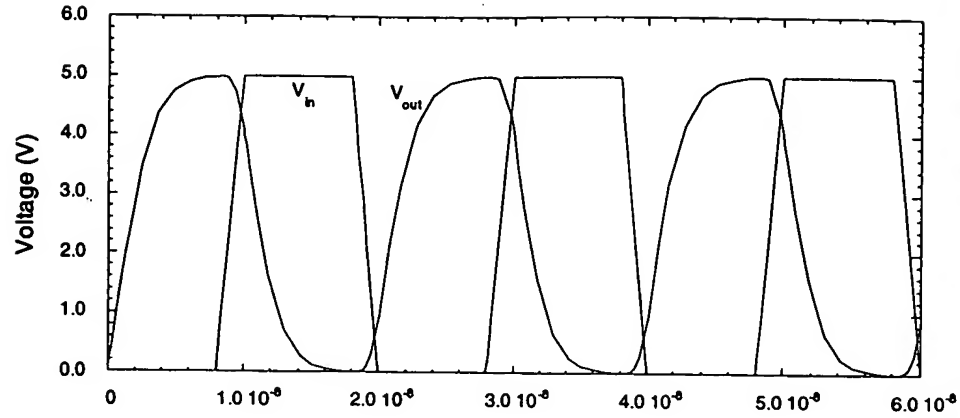
Example 6.6.

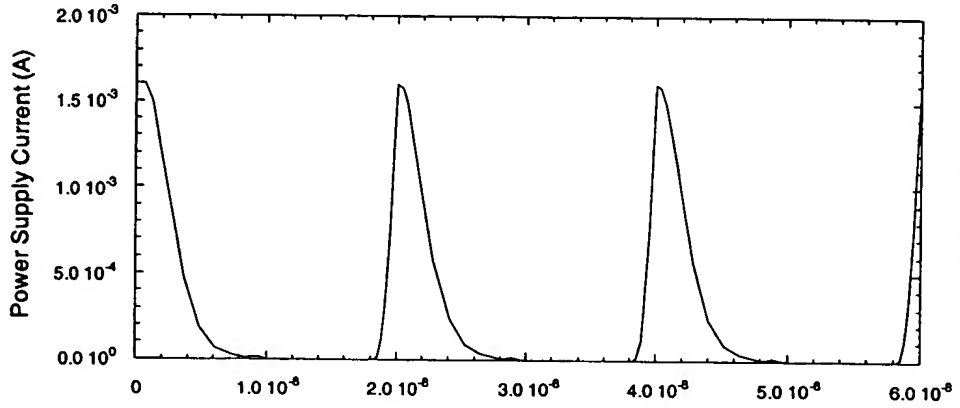
Consider the simple CMOS inverter circuit shown in Fig. 6.27. We will assume that the circuit is being driven by a square-wave input signal with period $T = 20$ ns, and that the total output load capacitance is equal to 1 pF. The power supply voltage is 5 V. Using the average dynamic power-dissipation formula (6.70) derived earlier, we can calculate the expected power dissipation to be $P_{avg} = 1.25$ mW.

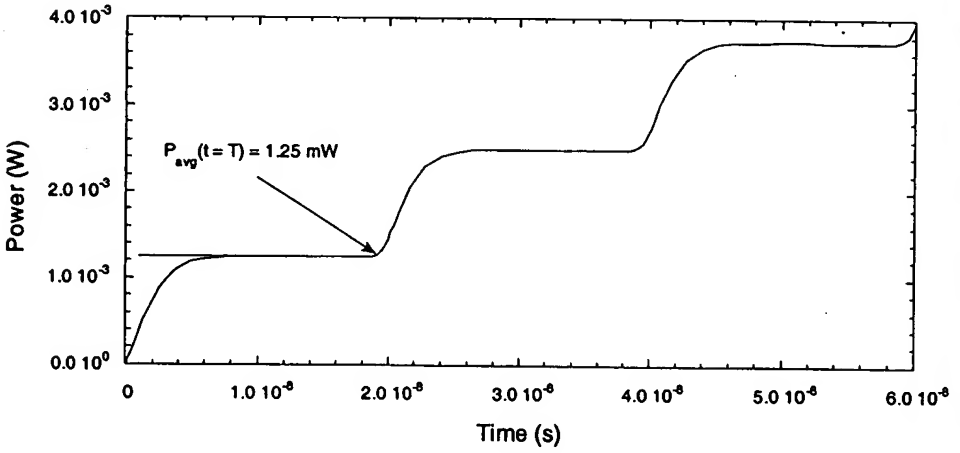
Now, the circuit with an attached power meter will be simulated using SPICE. The corresponding circuit input file is listed here for reference. The controlled current source coefficient is calculated as 0.025, according to (6.74). The resistance and capacitance values R_y and C_y are chosen as 100 kΩ and 100 pF to satisfy the condition $R_y C_y >> T$.

```
Power meter simulation:
mn 3 2 0 0 nmod w=10u l=1u
mp 3 2 4 1 pmod w=20u l=1u
vdd 1 0 5
vtstp 1 4 0
.model nmod nmos(vto=1 kp=20u)
.model pmod pmos(vto=-1 kp=10u)
vin 2 0 pulse(0 5 8n 2n 2n 8n 20n)
cl 3 0 1p
fp 0 9 vtstp 0.025
rp 9 0 100k
cp 9 0 100p
.tran 1n 60n uic
.print tran v(3) v(2)
.print tran i(vtstp)
.print tran v(9)
.end
```

The simulation results are plotted on the following page. It can be seen that here, the significant power supply current is being drawn from the voltage source V_{DD} only during the charge-up phase of the output capacitor. The power meter output voltage by the end of the first period corresponds to exactly 1.25 mW, as expected.


Voltage (V)
6.0
5.0
4.0
3.0
2.0
1.0
0.0
V_{in}
V_{out}
0
1.0 10⁻⁶
2.0 10⁻⁶
3.0 10⁻⁶
4.0 10⁻⁶
5.0 10⁻⁶
6.0 10⁻⁶


Power Supply Current (A)
2.0 10⁻³
1.5 10⁻³
1.0 10⁻³
5.0 10⁻⁴
0.0 10⁰
0
1.0 10⁻⁶
2.0 10⁻⁶
3.0 10⁻⁶
4.0 10⁻⁶
5.0 10⁻⁶
6.0 10⁻⁶


Power (W)
4.0 10⁻³
3.0 10⁻³
2.0 10⁻³
1.0 10⁻³
0.0 10⁰
$P_{avg}(t = T) = 1.25$ mW
0
1.0 10⁻⁶
2.0 10⁻⁶
3.0 10⁻⁶
4.0 10⁻⁶
5.0 10⁻⁶
6.0 10⁻⁶
Time (s)

Power-Delay Product

The power-delay product (PDP) is a fundamental parameter which is often used for measuring the quality and the performance of a CMOS process and gate design. As a physical quantity, the power-delay product can be interpreted as the average *energy* required for a gate to switch its output voltage from low to high and from high to low. We have already seen that in a CMOS logic gate, energy is dissipated (i) by the pMOS network while the output load capacitance C_{load} is being charged up from 0 to V_{DD}, and (ii) by the nMOS network while the output load capacitance is being charged down from V_{DD} to 0. Following a simple analysis procedure which is very similar to the one used for deriving the average dynamic power dissipation (6.69) in CMOS logic gates and ignoring the short-circuit and leakage currents, the amount of energy required to switch the output can be found as

$$PDP = C_{load}\, V_{DD}^{\,2} \tag{6.76}$$

The energy described by (6.76) is mainly dissipated as heat when the nMOS and pMOS transistors conduct current during switching. Thus, from a design point-of-view, it is desirable to minimize the power-delay product. Since the PDP is a function of the output load capacitance and the power supply voltage, the designer should try to keep both C_{load} and V_{DD} as small as possible when designing a CMOS logic gate. The power-delay product is also defined as

$$PDP = 2\, P^{*}_{avg}\, \tau_P \tag{6.77}$$

where P^{*}_{avg} is the average switching power dissipation at *maximum* operating frequency and τ_P is the average propagation delay, as defined in (6.4). The factor of 2 in (6.77) accounts for *two* transitions of the output, from low to high and from high to low. Using (6.69) and (6.4), this expression can be rewritten as

$$\begin{aligned} PDP &= 2\left(C_{load}\, V_{DD}^{\,2}\, f_{max}\right)\tau_P \\ &= 2\left[C_{load}\, V_{DD}^{\,2}\left(\frac{1}{\tau_{PHL}+\tau_{PLH}}\right)\right]\left(\frac{\tau_{PHL}+\tau_{PLH}}{2}\right) \\ &= C_{load}\, V_{DD}^{\,2} \end{aligned} \tag{6.78}$$

which is found to be identical to (6.76). Note that calculating PDP with the generic definition of P_{avg} (6.70) may result in a *misleading interpretation* that the amount of energy required per switching event is a function of the operating frequency.

References

1. N.H.E. Weste and K. Eshraghian, *Principles of CMOS VLSI Design–A Systems Perspective*, second edition, Reading, MA: Addison-Wesley, 1993.

2. J.P. Uyemura, *Fundamentals of MOS Digital Integrated Circuits*, Reading, MA: Addison-Wesley, 1988.

3. M. Shoji, *CMOS Digital Circuit Technology*, Englewood Cliffs, NJ: Prentice Hall, 1988.

4. C.P. Yuan and T.N. Trick, "A simple formula for the estimation of capacitance of two-dimensional interconnects in VLSI circuits," *IEEE Electron Device Letters*, vol. EDL-3, no. 12, pp. 391-393, December 1982.

5. D. Zhou, F.P. Preparata, and S. M. Kang, "Interconnection delay in very high-speed VLSI," *IEEE Transactions on Circuits and Systems*, vol. 38, no. 7, pp. 779-790, July 1991.

6. G. Bilardi, M. Pracchi, and F.P. Preparata, "A critique of network speed in VLSI models of computation," *IEEE Journal of Solid-State Circuits*, vol. DC-17, no. 4, pp. 696-702, August 1982

7. S.M. Kang, "Accurate simulation of power dissipation in VLSI circuits," *IEEE Journal of Solid-State Circuits*, vol. SC-21, no. 10, pp. 889-891, October 1986.

8. M. Horowitz and R.W. Dutton, "Resistance extraction from mask layout data," *IEEE Transactions on Computer-Aided Design*, vol. CAD-2, no. 3, pp. 145-150, July 1983.

9. A.E. Ruehli and P.A. Brennan, "Efficient capacitance calculations for three-dimensional multiconductor systems," *IEEE Transactions on Microwave Theory and Applications*, vol. MTT-21, no. 2, pp. 76-82, February 1973.

10. P. DeWilde, "New algebraic methods for modelling large-scale integrated circuits," *International Journal of Circuit Theory and Applications*, vol. 16, no. 4, pp. 473-503, October 1988.

11. T. Sakurai and A. R. Newton, "Alpha-power law MOSFET model and its application to CMOS inverter delay and other formulas," *IEEE Journal of Solid-State Circuits*, vol. 25, no. 2, pp. 584-594, April 1990.

12. H.B. Bakoglu, *Circuits, Interconnections and Packaging for VLSI*, Reading, MA: Addison-Wesley, 1990.

APPENDIX

Super Buffer Design

The term *super buffer* has been used to describe a chain of inverters designed to drive a large capacitive load with minimal signal propagation delay time. To reduce delay time, it is necessary for the buffer circuit to provide quickly a large amount of pull-up or pull-down current to charge or discharge the load capacitor. One seemingly obvious method would be to use large pMOS and nMOS transistors in the inverter driving the load capacitor. However, such a large buffer has a large input capacitance, which in turn creates a large load for the previous stage. Then an alert designer would suggest increasing the transistor sizes in the previous stage. If so, then what about the sizing of the transistors in the stage prior to the previous stage? Thus the effect of the large load can be propagated to many gates preceding the last-stage driver, and indeed such fine tuning of transistors is practiced in custom design. An alternative method of handling a large capacitive load is to use a super buffer between a logic gate facing the large load and the load itself as shown in Fig. A.1.

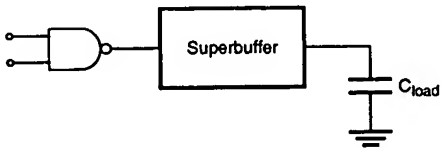


Figure A.1. Using a super buffer circuit to drive a large capacitive load.

Now a major objective of super buffer design becomes:

Given the load capacitance faced by a logic gate, design a scaled chain of N *inverters such that the delay time between the logic gate and the load capacitance node is minimized.*

To solve this problem, let us first introduce an equivalent inverter for the logic gate (NAND2 in this case). For simplicity, it is assumed that the pull-up and pull-down delays of the first-stage inverter driving an identical inverter are the same, say τ_0. The next design task is to determine the following:

- the number of stages, N
- the optimal scale factor, α

To determine these quantities, the following observations can be made under uniform α-scaling of inverters from one stage to the next in the super buffer shown in Fig. A.2.

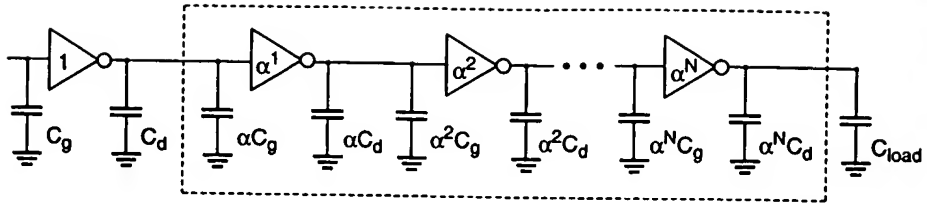


Figure A.2. Scaled super buffer circuit consisting of N inverter stages.

For the super buffer, the following observations can be made:

- C_g denotes the input capacitance of the first stage inverter
- C_d denotes the drain capacitance of the first stage inverter
- the inverters in the chain are scaled up by a factor of α per stage

- $C_{load} = \alpha^{N+1} C_g$ (A.1)
- all inverters have identical delay of $\tau_0 \left(C_d + \alpha C_g\right) / \left(C_d + C_g\right)$ (A.2)

where τ_0 represents the per-gate delay in the ring oscillator circuit with load capacitance $(C_d + C_g)$. Thus the total delay time from the input terminal to the load capacitance node becomes

$$\tau_{total} = (N+1)\tau_0 \left(\frac{C_d + \alpha C_g}{C_d + C_g} \right) \tag{A.3}$$

There are two unknowns in this equation. To solve for these unknowns, consider the relationship between α and N in (A.1), i.e.,

$$(N+1) = \frac{ln\left(\frac{C_{load}}{C_g}\right)}{ln\,\alpha} \tag{A.4}$$

Combining (A.3) and (A.4), the following delay relationship can be derived.

$$\tau_{total} = \frac{ln\left(\frac{C_{load}}{C_g}\right)}{ln\,\alpha} \tau_0 \left(\frac{C_d + \alpha C_g}{C_d + C_g} \right) \tag{A.5}$$

To minimize the delay, we set the derivative of (A.5) with respect to α equal to zero and solve for α.

$$\frac{\partial \tau_{total}}{\partial \alpha} = \tau_0 \ln\left(\frac{C_{load}}{C_g}\right)\left[-\frac{\frac{1}{\alpha}}{(\ln \alpha)^2}\left(\frac{C_d + \alpha C_g}{C_d + C_g}\right) + \frac{1}{\ln \alpha}\left(\frac{C_g}{C_d + C_g}\right)\right] = 0 \qquad \text{(A.6)}$$

Solving for α in (A.6) we obtain the following condition for the optimal scale factor.

$$\alpha(\ln \alpha - 1) = \frac{C_d}{C_g} \qquad \text{(A.7)}$$

A special case of the above equation occurs when the drain capacitance is neglected, i.e., $C_d = 0$. In that case, the optimal scale factor becomes the natural number $e = 2.718$. However, in reality the drain parasitics cannot be ignored and hence, (A.6) should be considered instead.

CHAPTER 7

COMBINATIONAL MOS LOGIC CIRCUITS

7.1. Introduction

Combinational logic circuits, or gates, which perform Boolean operations on multiple input variables and determine the outputs as Boolean functions of the inputs, are the basic building blocks of all digital systems. In this chapter, we will examine the static and dynamic characteristics of various combinational MOS logic circuits. It will be seen that many of the basic principles used in the design and analysis of MOS inverters in Chapters 5 and 6 can be directly applied to combinational logic circuits as well.

The first major class of combinational logic circuits to be presented in this chapter is the nMOS depletion-load gates. Our purpose for including nMOS depletion-load circuits here is mainly pedagogical, to emphasize the *load* concept, which is still being widely used in many areas in digital circuit design. We will examine simple circuit configurations such as two-input NAND and NOR gates and then expand our analysis to more general cases of multiple-input circuit structures. Next, the CMOS logic circuits will be presented in a similar fashion. We will stress the similarities and differences between the nMOS depletion-load logic and CMOS logic circuits and point out the advantages of CMOS gates with examples. The design of complex logic gates, which allows the realization of complex Boolean functions of multiple variables, will be examined in detail. Finally, we will devote the last section to CMOS transmission gates and to transmission gate (TG) logic circuits.

In its most general form, a combinational logic circuit, or gate, performing a Boolean function can be represented as a multiple-input single-output system, as depicted in Fig.

(a) Calculate the rise time and the fall time of the output signal using

(i) exact method (differential equations)
(ii) average current method

(b) Determine the *maximum* frequency of a periodic square-wave input signal so that the output voltage can still exhibit a full logic swing from 0 V to 5 V in each cycle.

(c) Calculate the dynamic power dissipation at this frequency.

(d) Assume that the output load capacitance is mainly dominated by fixed fan-out components (which are independent of W_n and W_p). We want to re-design the inverter so that the propagation delay times are reduced by 25%. Determine the required channel dimensions of the nMOS and the pMOS transistors. How does this re-design influence the switching (inversion) threshold?

7.1. All input variables are represented by node voltages, referenced to the ground potential. Using *positive logic convention*, the Boolean (or logic) value of "1" can be represented by a high voltage of V_{DD}, and the Boolean (or logic) value of "0" can be represented by a low voltage of 0. The output node is loaded with a capacitance C_L, which represents the combined parasitic device capacitances in the circuit and the interconnect capacitance components *seen* by the output node. This output load capacitance certainly plays a very significant role in the dynamic operation of the logic gate.

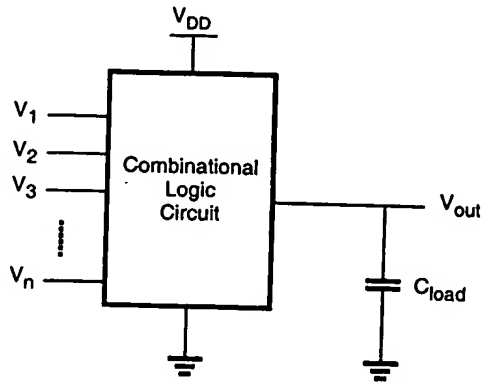


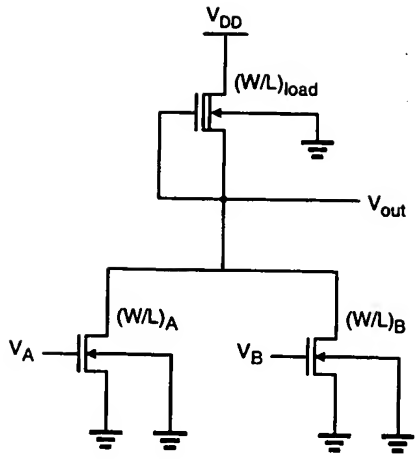
Figure 7.1. Generic combinational logic circuit (gate).

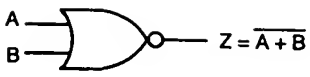
As in the simple inverter case, the voltage transfer characteristic (VTC) of a combinational logic gate provides valuable information on the DC operating performance of the circuit. Critical voltage points such as V_{OL} or V_{th} are considered to be important design parameters for combinational logic circuits. Other design parameters and concerns include the dynamic (transient) response characteristics of the circuit, the silicon area occupied by the circuit, and the amount of static and dynamic power dissipation.

7.2. MOS Logic Circuits with Depletion nMOS Loads

Two-Input NOR Gate

The first circuit to be examined in this section is the two-input NOR gate. The circuit diagram, the logic symbol, and the corresponding truth table of the gate are given in Fig. 7.2. The Boolean OR operation is performed by the parallel connection of the two enhancement-type nMOS driver transistors. If the input voltage V_A *or* the input voltage V_B is equal to the logic-high level, the corresponding driver transistor turns on and provides a conducting path between the output node and the ground. Hence, the output voltage becomes low. In this case, the circuit operates like a depletion-load inverter with respect to its static behavior. A similar result is achieved when both V_A and V_B are high, in which case two parallel conducting paths are created between the output node and the ground. If, on the other hand, both V_A and V_B are low, both driver transistors remain cut off. The output node voltage is pulled to a logic-high level by the depletion-type nMOS load transistor.





V_A	V_B	V_{out}
low	low	high
low	high	low
high	low	low
high	high	lower

Figure 7.2. A two-input depletion-load NOR gate, its logic symbol, and the corresponding truth table. Note that the substrates of all transistors are connected to ground.

The DC analysis of the circuit can be simplified significantly by considering the structural similarities between this circuit and the simple nMOS depletion-load inverter. In the following, the calculation of output low and output high voltages will be examined.

Calculation of V_{OH}

When both input voltages V_A and V_B are lower than the corresponding driver threshold voltage, the driver transistors are turned off and conduct no drain current. Consequently, the load device, which operates in the linear region, also has zero drain current. In particular, its linear region current equation becomes

$$I_{D,load} = \frac{k_{n,load}}{2} \cdot \left[2\left|V_{T,load}(V_{OH})\right| \cdot (V_{DD} - V_{OH}) - (V_{DD} - V_{OH})^2 \right] = 0 \qquad (7.1)$$

The solution of this equation gives $V_{OH} = V_{DD}$.

Calculation of V_{OL}

To calculate the output low voltage V_{OL}, we must consider three different cases, i.e., three different input voltage combinations, which produce a conducting path from the output node to the ground. These cases are

(i) $V_A = V_{OH}$ $V_B = V_{OL}$
(ii) $V_A = V_{OL}$ $V_B = V_{OH}$
(iii) $V_A = V_{OH}$ $V_B = V_{OH}$

For the first two cases, (i) and (ii), the NOR circuit reduces to a simple nMOS depletion-load inverter. Assuming that the threshold voltages of the two enhancement-type driver transistors are identical ($V_{T0,A} = V_{T0,B} = V_{T0}$), the driver-to-load ratio of the corresponding inverter can be found as follows. In case (i), where the driver transistor A is on, the ratio is

$$k_R = \frac{k_{driver,A}}{k_{load}} = \frac{k'_{n,driver}\left(\frac{W}{L}\right)_A}{k'_{n,load}\left(\frac{W}{L}\right)_{load}} \tag{7.2}$$

In case (ii), where the driver transistor B is on, the ratio is

$$k_R = \frac{k_{driver,B}}{k_{load}} = \frac{k'_{n,driver}\left(\frac{W}{L}\right)_B}{k'_{n,load}\left(\frac{W}{L}\right)_{load}} \tag{7.3}$$

The output low voltage level V_{OL} in both cases is found by using (5.54), as follows:

$$V_{OL} = V_{OH} - V_{T0} - \sqrt{(V_{OH} - V_{T0})^2 - \left(\frac{k_{load}}{k_{driver}}\right) \cdot \left|V_{T,load}(V_{OL})\right|^2} \tag{7.4}$$

Note that if the (W/L) ratios of both drivers are identical, i.e., $(W/L)_A = (W/L)_B$, the output low voltage (V_{OL}) values calculated for case (i) and case (ii) will be identical.

In case (iii), where both driver transistors are turned on, the saturated load current $I_{D,load}$ is the sum of the two linear-mode driver currents.

$$I_{D,load} = I_{D,driverA} + I_{D,driverB} \tag{7.5}$$

$$\frac{k_{load}}{2}\left|V_{T,load}(V_{OL})\right|^2 = \frac{k_{driver,A}}{2}\left[2(V_A - V_{T0})V_{OL} - V_{OL}^2\right] + \frac{k_{driver,B}}{2}\left[2(V_B - V_{T0})V_{OL} - V_{OL}^2\right] \tag{7.6}$$

Since the gate voltages of both driver transistors are equal ($V_A = V_B = V_{OH}$), we can devise an *equivalent* driver-to-load ratio for the NOR structure:

$$k_R = \frac{k_{driver,A} + k_{driver,B}}{k_{load}} = \frac{k'_{n,driver}\left[\left(\frac{W}{L}\right)_A + \left(\frac{W}{L}\right)_B\right]}{k'_{n,load}\left(\frac{W}{L}\right)_{load}} \tag{7.7}$$

Thus, the NOR gate with both of its inputs tied to a logic-high voltage is replaced with an nMOS depletion-load inverter circuit with the driver-to-load ratio given by (7.7). The output voltage level in this case is

$$V_{OL} = V_{OH} - V_{T0} - \sqrt{(V_{OH} - V_{T0})^2 - \left(\frac{k_{load}}{k_{driver,A} + k_{driver,B}}\right) \cdot |V_{T,load}(V_{OL})|^2} \tag{7.8}$$

Note that the V_{OL} given by (7.8) is *lower* than the V_{OL} values calculated for case (i) and for case (ii), when only one input is logic-high. We conclude that the worst-case condition from the static operation viewpoint, i.e., the highest possible V_{OL} value, is observed in case (i) or in (ii).

This result also suggests a simple design strategy for NOR gates. Usually, we have to achieve a certain maximum V_{OL} for the worst case, i.e., when only one input is high. Thus, we assume that one input (either V_A or V_B) is logic-high and determine the driver-to-load ratio of the resulting inverter using (7.4). Then set

$$k_{driver,A} = k_{driver,B} = k_R\, k_{load} \tag{7.9}$$

This design choice yields two identical driver transistors, which guarantee the required value of V_{OL} in the worst case. When both inputs are logic-high, the output voltage is even lower than the required maximum V_{OL}, thus the design constraint is satisfied.

Exercise 7.1.

Consider the depletion-load nMOS NOR2 gate shown in Fig. 7.2, with the following parameters: $\mu_n C_{ox} = 25\ \mu A/V^2$, $V_{T0,driver} = 1.0$ V, $V_{T0,load} = -3.0$ V, $\gamma = 0.4\ V^{1/2}$, and $|2\phi_F| = 0.6$ V. The transistor dimensions are given as $(W/L)_A = 2$, $(W/L)_B = 4$, and $(W/L)_{load} = 1/3$. The power supply voltage is $V_{DD} = 5$ V.

Calculate the output voltage levels for all four valid input voltage combinations.

Generalized NOR Structure with Multiple Inputs

At this point, we can expand our analysis to generalized n-input NOR gates, which consist of n parallel driver transistors, as shown in Fig. 7.3. Note that the combined current I_D in this circuit is supplied by the driver transistors which are turned on, i.e., transistors which have gate voltages higher than the threshold voltage V_{T0}.

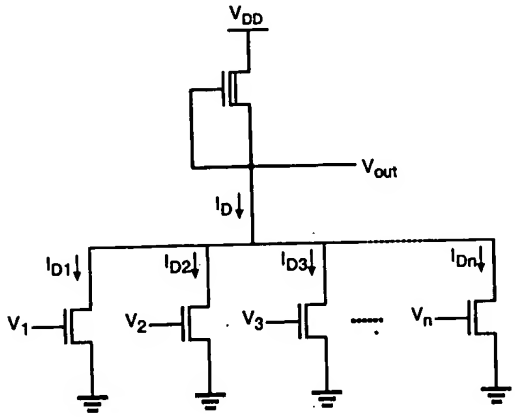


Figure 7.3. Generalized n-input NOR gate.

The combined pull-down current can then be expressed as follows:

$$I_D = \sum_{k(on)} I_{D,k} = \begin{cases} \displaystyle\sum_{k(on)} \frac{\mu_n C_{ox}}{2}\left(\frac{W}{L}\right)_k \left[2\left(V_{GS,k} - V_{T0}\right)V_{out} - V_{out}^2\right] & \textit{linear} \\ \displaystyle\sum_{k(on)} \frac{\mu_n C_{ox}}{2}\left(\frac{W}{L}\right)_k \left(V_{GS,k} - V_{T0}\right)^2 & \textit{saturation} \end{cases} \tag{7.10}$$

Assuming that the input voltages of all driver transistors are identical,

$$V_{GS,k} = V_{GS} \qquad \textit{for} \qquad k = 1, 2, \ldots, n \tag{7.11}$$

the pull-down current expression can be rewritten as

$$I_D = \begin{cases} \dfrac{\mu_n C_{ox}}{2}\left(\displaystyle\sum_{k(on)}\left(\frac{W}{L}\right)_k\right)\left[2\left(V_{GS} - V_{T0}\right)V_{out} - V_{out}^2\right] & \textit{linear} \\ \dfrac{\mu_n C_{ox}}{2}\left(\displaystyle\sum_{k(on)}\left(\frac{W}{L}\right)_k\right)\left(V_{GS} - V_{T0}\right)^2 & \textit{saturation} \end{cases} \tag{7.12}$$

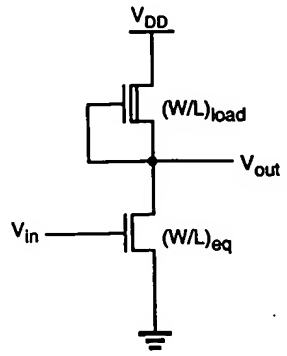


Figure 7.4. Equivalent inverter circuit corresponding to the n-input NOR gate.

Thus, the multiple-input NOR gate can also be reduced to an equivalent inverter, shown in Fig. 7.4, for static analysis. The (W/L) ratio of the driver transistor here is

$$\left(\frac{W}{L}\right)_{equivalent} = \sum_{k(on)} \left(\frac{W}{L}\right)_k \tag{7.13}$$

Note that the source terminals of all enhancement-type nMOS driver transistors in the NOR gate are connected to ground. Thus, the drivers do not experience any substrate-bias effect. The depletion-type nMOS load transistor, however, is subject to substrate-bias effect, since its source is connected to the output node, and its source-to-substrate voltage is $V_{SB} = V_{out}$.

Transient Analysis of NOR Gate

Figure 7.5 shows the two-input NOR (NOR2) gate with all of its relevant parasitic device capacitances. As in the inverter case, we can combine the capacitances seen in Fig. 7.5 into one lumped capacitance, connected between the output node and the ground. The value of this combined load capacitance, C_{load}, can be found as

$$C_{load} = C_{gd,A} + C_{gd,B} + C_{gd,load} + C_{db,A} + C_{db,B} + C_{sb,load} + C_{wire} \tag{7.14}$$

Note that the output load capacitance given in (7.14) is valid for simultaneous as well as for single-input switching, i.e., the load capacitance C_{load} will be present at the output node even if only one input is active and all other inputs are low. This fact must be taken into account in calculations using the inverter equivalent of the NOR gate. The load capacitance at the output node of the *equivalent* inverter corresponding to a NOR gate is always *larger* than the total lumped load capacitance of an actual inverter with the same dimensions. Hence, while the static (DC) behaviors of the NOR gate and the inverter are

essentially equivalent in this case, the actual transient response of the NOR gate will be slower than that of the inverter.

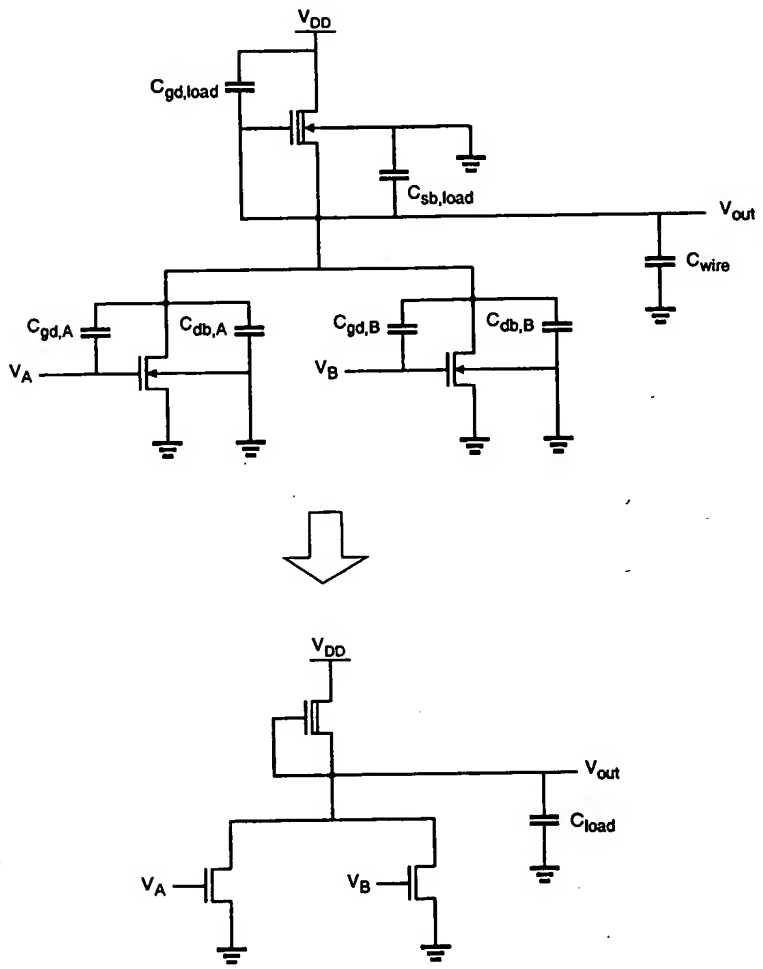


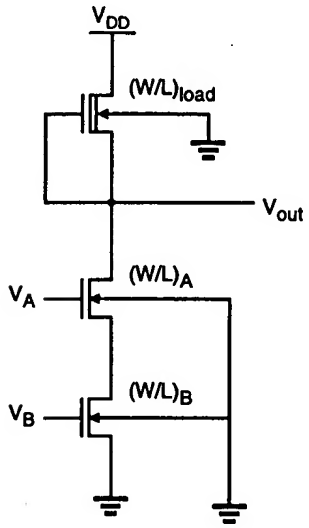
Figure 7.5. Parasitic device capacitances in the NOR2 gate and the lumped equivalent load capacitance. The gate-to-source capacitances of the driver transistors are included in the load of the previous stages driving the inputs A and B.

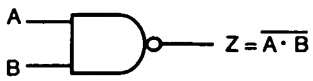
Two-Input NAND Gate

Next, we will examine the two-input NAND (NAND2) gate. The circuit diagram, the logic symbol, and the corresponding truth table of the gate are given in Fig. 7.6. The Boolean AND operation is performed by the series connection of the two enhancement-type nMOS driver transistors. There is a conducting path between the output node and the

ground only if the input voltage V_A *and* the input voltage V_B are equal to logic-high, i.e., only if both of the series-connected drivers are turned on. In this case, the output voltage will be low, which is the complemented result of the AND operation. Otherwise, either one or both of the driver transistors will be off, and the output voltage will be pulled to a logic-high level by the depletion-type nMOS load transistor.

Figure 7.6 shows that all transistors except the one closest to the ground are subject to substrate-bias effect, since their source voltages are larger than zero. We have to consider this fact in detailed calculations. For all of the three input combinations which produce a logic-high output voltage, the corresponding V_{OH} value can easily be found as $V_{OH} = V_{DD}$. The calculation of the logic-low voltage V_{OL}, on the other hand, requires a closer investigation.





V_A	V_B	V_{out}
low	low	high
low	high	high
high	low	high
high	high	low

Figure 7.6. A two-input depletion-load NAND gate, its logic symbol, and the corresponding truth table. Notice the substrate-bias effect for all nMOS transistors except one.

Consider the NAND2 gate with both of its inputs equal to V_{OH}, as shown in Fig. 7.7. It can easily be seen that the drain currents of all transistors in the circuit are equal to each other.

$$I_{D,load} = I_{D,driverA} = I_{D,driverB} \tag{7.15}$$

$$\begin{aligned} \frac{k_{load}}{2}\left|V_{T,load}(V_{OL})\right|^2 &= \frac{k_{driver,A}}{2}\left[2\left(V_{GS,A}-V_{T,A}\right)V_{DS,A}-V_{DS,A}^2\right] \\ &= \frac{k_{driver,B}}{2}\left[2\left(V_{GS,B}-V_{T,B}\right)V_{DS,B}-V_{DS,B}^2\right] \end{aligned} \tag{7.16}$$

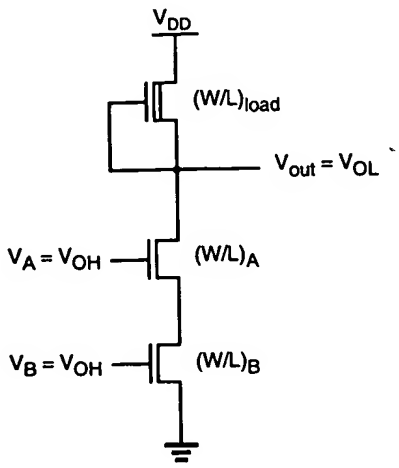


Figure 7.7. The NAND2 gate with both of its inputs at logic-high level.

The gate-to-source voltages of both driver transistors can be assumed to be approximately equal to V_{OH}. Also, we may neglect, for simplicity, the substrate-bias effect for driver transistor A, and assume $V_{T,A} = V_{T,B} = V_{T0}$, since the source-to-substrate voltage of driver A is relatively low. The drain-to-source voltages of both driver transistors can then be solved from (7.16) as

$$V_{DS,A} = V_{OH} - V_{T0} - \sqrt{\left(V_{OH} - V_{T0}\right)^2 - \left(\frac{k_{load}}{k_{driver,A}}\right) \cdot \left|V_{T,load}\left(V_{OL}\right)\right|^2} \tag{7.17}$$

$$V_{DS,B} = V_{OH} - V_{T0} - \sqrt{\left(V_{OH} - V_{T0}\right)^2 - \left(\frac{k_{load}}{k_{driver,B}}\right) \cdot \left|V_{T,load}\left(V_{OL}\right)\right|^2} \tag{7.18}$$

Let the two driver transistors be identical, i.e., $k_{driver,A} = k_{driver,B} = k_{driver}$. Noting that the output voltage V_{OL} is equal to the sum of the drain-to-source voltages of both drivers, we obtain

$$V_{OL} \approx 2\left(V_{OH} - V_{T0} - \sqrt{\left(V_{OH} - V_{T0}\right)^2 - \left(\frac{k_{load}}{k_{driver}}\right) \cdot \left|V_{T,load}\left(V_{OL}\right)\right|^2} \right) \tag{7.19}$$

The following analysis gives a better and more accurate view of the operation of two series-connected driver transistors. Consider the two identical enhancement-type nMOS transistors with their gate terminals connected. At this point, the only simplifying

assumption will be $V_{T,A} = V_{T,B} = V_{T0}$. When both driver transistors are in the linear region, the drain currents can be written as

$$I_{D,A} = \frac{k_{driver}}{2}\left[2\left(V_{GS,A} - V_{T0}\right)V_{DS,A} - V_{DS,A}^2\right] \tag{7.20}$$

$$I_{D,B} = \frac{k_{driver}}{2}\left[2\left(V_{GS,B} - V_{T0}\right)V_{DS,B} - V_{DS,B}^2\right] \tag{7.21}$$

Since $I_{D,A} = I_{D,B}$, this current can also be expressed as

$$I_D = I_{D,A} = I_{D,B} = \frac{I_{D,A} + I_{D,B}}{2} \tag{7.22}$$

Using $V_{GS,A} = V_{GS,B} - V_{DS,B}$, (7.22) yields

$$I_D = \frac{k_{driver}}{4}\left[2\left(V_{GS,B} - V_{T0}\right)\left(V_{DS,A} + V_{DS,B}\right) - \left(V_{DS,A} + V_{DS,B}\right)^2\right] \tag{7.23}$$

Now let $V_{GS} = V_{GS,B}$ and $V_{DS} = V_{DS,A} + V_{DS,B}$. The drain-current expression can then be written as follows.

$$I_D = \frac{k_{driver}}{4}\left[2\left(V_{GS} - V_{T0}\right)V_{DS} - V_{DS}^2\right] \tag{7.24}$$

Thus, two nMOS transistors connected in series and with the same gate voltage behave like *one* nMOS transistor with $k_{eq} = 0.5\, k_{driver}$.

Generalized NAND Structure with Multiple Inputs

At this point, we expand our analysis to generalized n-input NAND gates, which consist of n series-connected driver transistors, as shown in Fig. 7.8. Neglecting the substrate-bias effect, and assuming that the threshold voltages of all transistors are equal to V_{T0}, the driver current I_D in the linear region can be derived as in Eq. (7.25) whereas I_D in saturation is taken as its extension.

$$I_D = \frac{\mu_n C_{ox}}{2}\left(\frac{1}{\sum_{k(on)}\frac{1}{\left(\frac{W}{L}\right)_k}}\right)\cdot\begin{cases}\left[2\left(V_{in} - V_{T0}\right)V_{out} - V_{out}^2\right] & \textit{linear}\\ \left(V_{in} - V_{T0}\right)^2 & \textit{saturation}\end{cases} \tag{7.25}$$

Hence, the (W/L) ratio of the equivalent driver transistor is

$$\left(\frac{W}{L}\right)_{equivalent} = \frac{1}{\sum_{k(on)} \frac{1}{\left(\frac{W}{L}\right)_k}} \tag{7.26}$$

If the series-connected transistors are identical, i.e., $(W/L)_1 = (W/L)_2 = \ldots = (W/L)$, the width-to-length ratio of the equivalent transistor becomes

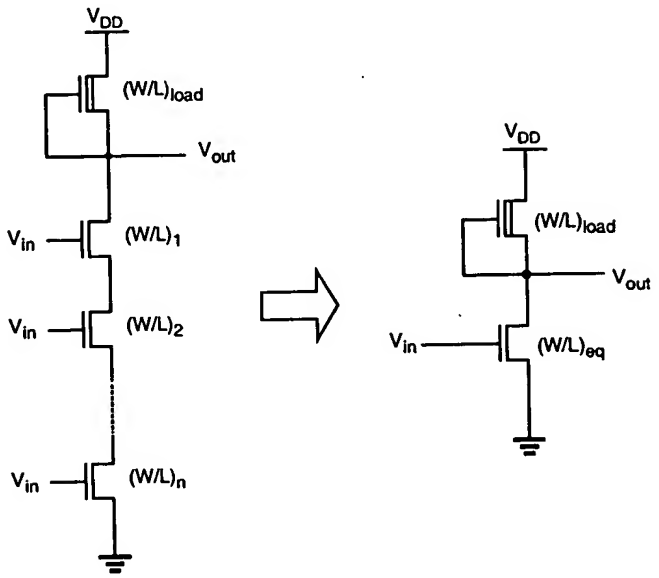


Figure 7.8. The generalized NAND structure and its inverter equivalent.

$$\left(\frac{W}{L}\right)_{equivalent} = \frac{1}{n}\left(\frac{W}{L}\right) \tag{7.27}$$

The NAND design strategy which emerges from this analysis is summarized as follows, for an n-input NAND. First, we determine the (W/L) ratios for an *equivalent inverter* that satisfies the required V_{OL} value. This gives us the driver transistor ratio $(W/L)_{driver}$ and the load transistor ratio $(W/L)_{load}$. Then, we set the (W/L) ratios of all NAND driver transistors as $(W/L)_1 = (W/L)_2 = \ldots = n\,(W/L)_{driver}$. This guarantees that the series structure consisting of n driver transistors has an equivalent (W/L) ratio of $(W/L)_{driver}$ when all inputs are logic-high.

For a two-input NAND gate, this means that each driver transistor must have a (*W/L*) ratio twice that of the equivalent inverter driver. If the area occupied by the depletion-type load transistor is negligible, the resulting NAND2 structure will occupy approximately four times the area occupied by the equivalent inverter which has the same static characteristics.

Transient Analysis of NAND Gate

Figure 7.9 shows a NAND2 gate with all parasitic device capacitances. As in the inverter case, we can combine the capacitances seen in Fig. 7.9 into one lumped capacitance, connected between the output node and the ground. The value of the lumped capacitance C_{load}, however, depends on the input voltage conditions.

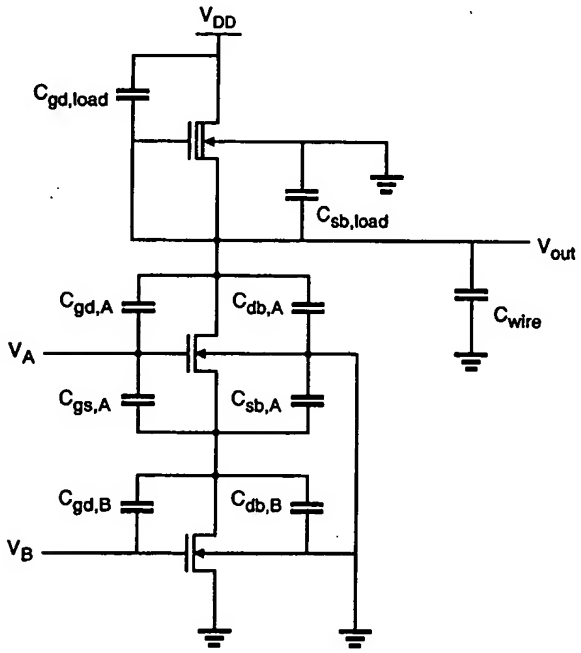


Figure 7.9. Parasitic device capacitances in the NAND2 gate.

Assume, for example, that the input V_A is equal to V_{OH} and the other input V_B is switching from V_{OH} to V_{OL}. In this case, both the output voltage V_{out} and the internal node voltage V_x will rise, resulting in

$$C_{load} = C_{gd,load} + C_{gd,A} + C_{gd,B} + C_{gs,A} + C_{db,A} + C_{db,B} + C_{sb,A} + C_{sb,load} + C_{wire} \tag{7.28}$$

Note that this value is quite conservative and fully reflects the internal node capacitances into the lumped output capacitance C_{load}. In reality, only a fraction of the internal node capacitance is reflected into C_{load}.

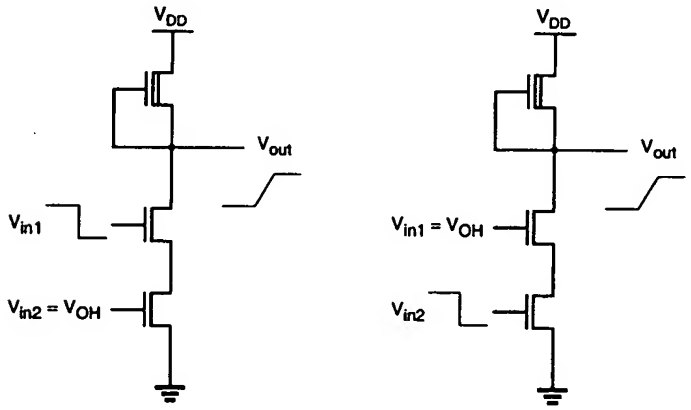
Now consider another case where V_B is equal to V_{OH} and V_A switches from V_{OH} to V_{OL}. In this case, the output voltage V_{out} will rise, but the internal node voltage V_x will remain low because the bottom driver transistor is on. Thus, the lumped output capacitance is

$$C_{load} = C_{gd,load} + C_{gd,A} + C_{db,A} + C_{sb,load} + C_{wire} \tag{7.29}$$

It should be noted that the load capacitance in this case is smaller than the load capacitance found in the previous case. Thus, it is expected that the high-to-low switching delay from signal B connected to the bottom transistor is larger than the high-to-low switching delay from signal A connected to the top transistor.

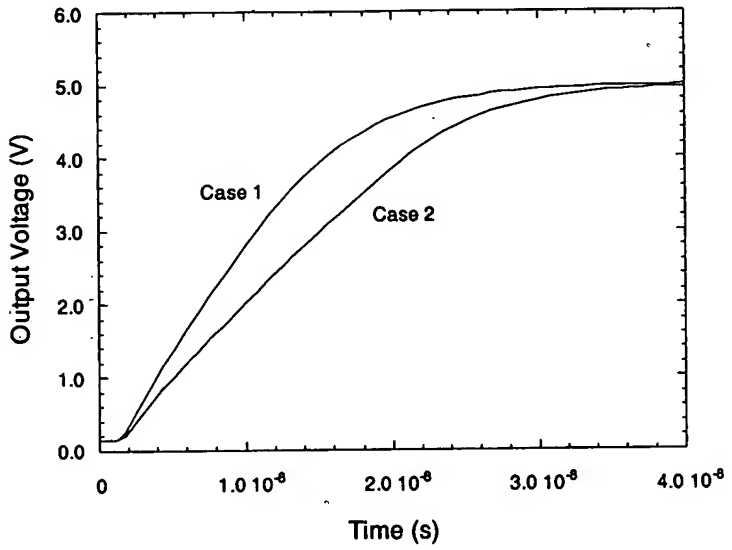
Example 7.1.

A depletion-load nMOS NAND2 gate is simulated with SPICE for the two different input switching events described above. The SPICE input file of the circuit is listed in the following. Note that the total capacitance between the intermediate node X and the ground is assumed to be half of the total capacitance appearing between the output node and the ground.



Case 1 Case 2

```
NAND2 circuit delay analysis
m1 3 1 0 0 mn w=5u l=1u
m2 4 2 3 0 mn w=5u l=1u
m3 5 4 4 0 mnd w=1u l=3u
c1 4 0 0.1p
cp 3 0 0.05p
vdd 5 0 dc 5.0
* case 1 (upper input switching from high to low)
vin1 2 0 dc pulse (5.0 0.0 1ns 1ns 2ns 40ns 50ns)
vin2 1 0 dc 5.0
* case 2 (lower input switching from high to low)
* vin1 2 0 dc 5.0
* vin2 1 0 dc pulse (5.0 0.0 1ns 1ns 2ns 40ns 50ns)
.model mn nmos (vto=1.0 kp=25u gamma=0.4)
.model mnd nmos (vto=-3.0 kp=25u gamma=0.4)
.tran 0.1ns 40ns
.print tran v(1) v(2) v(4)
.end
```



The simulated transient response of the NAND2 gate for both cases is plotted against time above. The time delay difference between the two cases is clearly visible. In fact, the propagation delay time in Case 2 is about 30% larger than that in Case 1, which proves that the input switching order has a significant influence on speed.

CMOS NOR2 (Two-Input NOR) Gate

The design and analysis of CMOS combinational logic circuits can be based on the basic principles developed for the nMOS depletion-load logic circuits in the previous section. Figure 7.10 shows the circuit diagram of a two-input CMOS NOR gate. Note that the circuit consists of a parallel-connected n-net and a series-connected complementary p-net. The input voltages V_A and V_B are applied to the gates of one nMOS and one pMOS transistor.

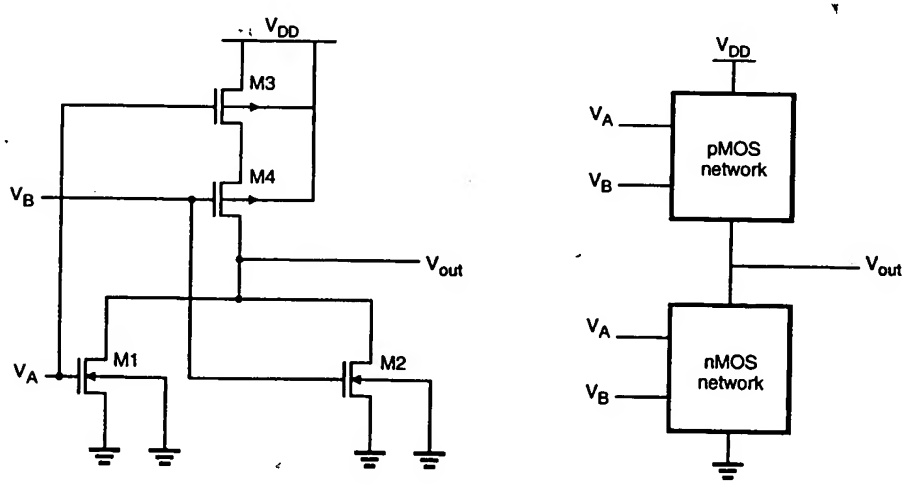


Figure 7.10. A CMOS NOR2 gate and its complementary operation: Either the nMOS network is on and the pMOS network is off, or the pMOS network is on and the nMOS network is off.

The complementary nature of the operation can be summarized as follows: When either one or both inputs are high, i.e., when the n-net creates a conducting path between the output node and the ground, the p-net is cut-off. On the other hand, if both input voltages are low, i.e., the n-net is cut-off, then the p-net creates a conducting path between the output node and the supply voltage V_{DD}. Thus, the dual or complementary circuit structure allows that, for any given input combination, the output is connected either to V_{DD} or to ground via a low-resistance path. A DC current path between the V_{DD} and ground is not established for any of the input combinations. This results in the fully complementary operation mode already examined for the simple CMOS inverter circuit.

The output voltage of the CMOS NOR2 gate will attain a logic-low voltage of $V_{OL} = 0$ and a logic-high voltage of $V_{OH} = V_{DD}$. For circuit design purposes, the switching threshold voltage V_{th} of the CMOS gate emerges as an important design criterion. We start our analysis of the switching threshold by assuming that both input voltages switch simultaneously, i.e., $V_A = V_B$. Furthermore, it is assumed that the device sizes in each block are identical, $(W/L)_{n,A} = (W/L)_{n,B}$ and $(W/L)_{p,A} = (W/L)_{p,B}$, and the substrate-bias effect for the pMOS transistors is neglected for simplicity.

By definition, the output voltage is equal to the input voltage at the switching threshold.

$$V_A = V_B = V_{out} = V_{th} \tag{7.30}$$

It is obvious that the two parallel nMOS transistors are saturated at this point, because $V_{GS} = V_{DS}$. The combined drain current of the two nMOS transistors is

$$I_D = k_n \left(V_{th} - V_{T,n}\right)^2 \tag{7.31}$$

Thus, we obtain the first equation for the switching threshold V_{th}.

$$V_{th} = V_{T,n} + \sqrt{\frac{I_D}{k_n}} \tag{7.32}$$

Examination of the p-net in Fig. 7.10 shows that the pMOS transistor M3 operates in the linear region, while the other pMOS transistor, M4, is in saturation for $V_{in} = V_{out}$. Thus,

$$I_{D3} = \frac{k_p}{2}\left[2\left(V_{DD} - V_{th} - \left|V_{T,p}\right|\right)V_{SD3} - V_{SD3}^2\right] \tag{7.33}$$

$$I_{D4} = \frac{k_p}{2}\left(V_{DD} - V_{th} - \left|V_{T,p}\right| - V_{SD3}\right)^2 \tag{7.34}$$

The drain currents of both pMOS transistors are identical, i.e., $I_{D3} = I_{D4} = I_D$. Thus,

$$V_{DD} - V_{th} - \left|V_{T,p}\right| = 2\sqrt{\frac{I_D}{k_p}} \tag{7.35}$$

This yields the second equation of the switching threshold voltage V_{th}. Combining (7.32) and (7.35), we obtain

$$V_{th}(\text{NOR2}) = \frac{V_{T,n} + \frac{1}{2}\sqrt{\frac{k_p}{k_n}}\left(V_{DD} - \left|V_{T,p}\right|\right)}{1 + \frac{1}{2}\sqrt{\frac{k_p}{k_n}}} \tag{7.36}$$

Now compare this expression with the switching threshold voltage of the CMOS inverter, which was derived in Chapter 5.

$$V_{th}(\text{INR}) = \frac{V_{T,n} + \sqrt{\frac{k_p}{k_n}}\left(V_{DD} - |V_{T,p}|\right)}{1 + \sqrt{\frac{k_p}{k_n}}} \tag{7.37}$$

If $k_n = k_p$ and $V_{T,n} = |V_{T,p}|$, the switching threshold of the CMOS inverter is equal to $V_{DD}/2$. Using the same parameters, the switching threshold of the NOR2 gate is

$$V_{th}(\text{NOR2}) = \frac{V_{DD} + V_{T,n}}{3} \tag{7.38}$$

which is not equal to $V_{DD}/2$. For example, when $V_{DD} = 5$ V and $V_{T,n} = |V_{T,p}| = 1$ V, the switching threshold voltages of the NOR2 gate and the inverter are

$$V_{th}(\text{NOR2}) = 2 \text{ V}$$

$$V_{th}(\text{INR}) = 2.5 \text{ V}$$

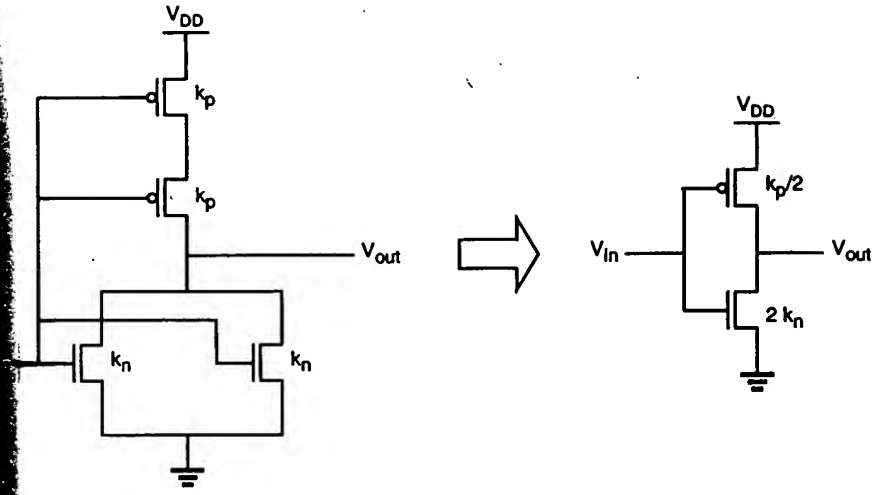
The switching threshold voltage of the NOR2 gate can also be obtained by using the equivalent-inverter approach. When both inputs are identical, the parallel-connected nMOS transistors can be represented by a single nMOS transistor with $2k_n$. Similarly, the series-connected pMOS transistors are represented by a single pMOS transistor with $k_p/2$. The resulting equivalent CMOS inverter is shown in Fig. 7.11.

Using the inverter switching threshold expression (7.37) for the equivalent inverter circuit, we obtain

$$V_{th}(\text{NOR2}) = \frac{V_{T,n} + \sqrt{\frac{k_p}{4k_n}}\left(V_{DD} - |V_{T,p}|\right)}{1 + \sqrt{\frac{k_p}{4k_n}}} \tag{7.39}$$

which is identical to (7.36).

From (7.36), we can easily derive simple design guidelines for the NOR2 gate. For example, in order to achieve a switching threshold voltage of $V_{DD}/2$ for simultaneous switching, we have to set $V_{T,n} = |V_{T,p}|$ and $k_p = 4\,k_n$.



...11. A CMOS NOR2 gate and its inverter equivalent.

...re 7.12 shows the CMOS NOR2 gate with the parasitic device capacitances, the ...equivalent, and the corresponding lumped output load capacitance. In the worst ...total lumped load capacitance is assumed to be equal to the sum of all internal ...device capacitances seen in Fig. 7.12.

...AND2 (Two-Input NAND) Gate

...13 shows a two-input CMOS NAND (NAND2) gate. The operating principle of ...it is the exact dual of the CMOS NOR2 operation examined earlier. The n-net ...of two series-connected nMOS transistors creates a conducting path between ...node and the ground only if both input voltages are logic-high, i.e., are equal ...this case, both of the parallel-connected pMOS transistors in the p-net will be ...other input combinations, either one or both of the pMOS transistors will be ..., while the n-net is cut-off, thus creating a current path between the output node ...power supply voltage.

...an analysis similar to the one developed for the NOR2 gate, we can easily ...the switching threshold for the CMOS NAND2 gate. Again, we will assume that ...sizes in each block are identical, with $(W/L)_{n,A} = (W/L)_{n,B}$ and $(W/L)_{p,A} =$ ... The switching threshold for this gate is then found as

$$V_M(\text{NAND2}) = \frac{V_{T,n} + 2\sqrt{\frac{k_p}{k_n}}\left(V_{DD} - |V_{T,p}|\right)}{1 + 2\sqrt{\frac{k_p}{k_n}}} \tag{7.40}$$

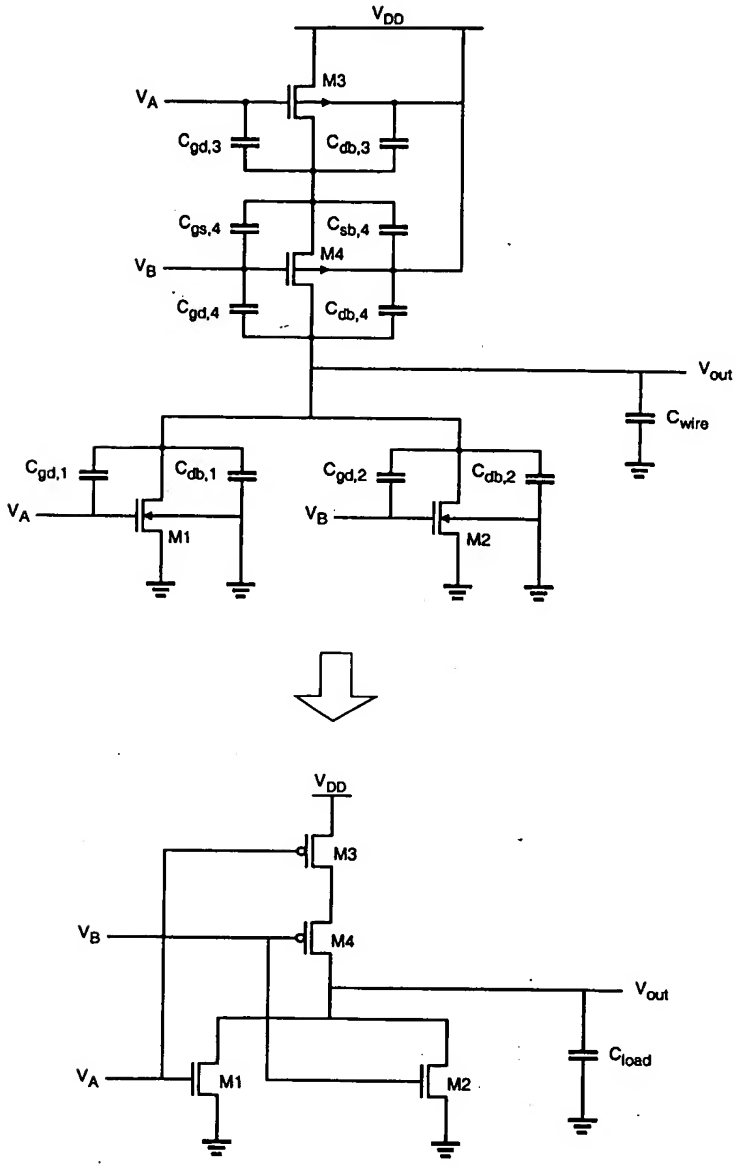


Figure 7.12. Parasitic device capacitances of the CMOS NOR2 circuit and the simplified equivalent with the lumped output load capacitance.

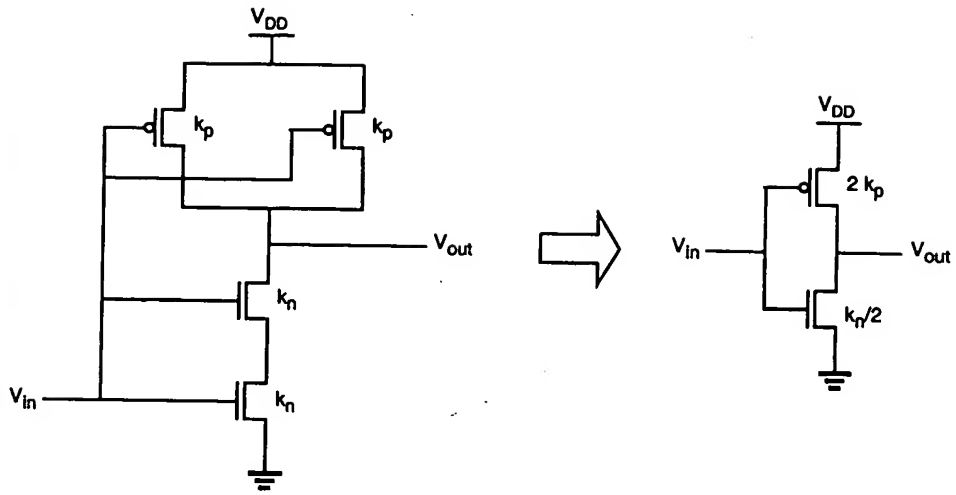


Figure 7.13. A CMOS NAND2 gate and its inverter equivalent.

As we can see from (7.40), a switching threshold voltage of $V_{DD}/2$ (for simultaneous switching) is achieved by setting $V_{T,n} = |V_{T,p}|$ and $k_n = 4\ k_p$ in the NAND2.

At this point, we can state the following observation about the area requirements of CMOS combinational logic gates. In comparison with equivalent nMOS depletion-load logic, the total number of transistors in CMOS gates is about twice the number of transistors in nMOS gates ($2n$ vs. $(n+1)$ for n inputs). The silicon area occupied by the CMOS gate, however, is not necessarily twice the area occupied by the nMOS depletion-load gate, since a significant portion of the silicon area must be reserved for signal routing and contacts in both cases. Thus, the area disadvantage of CMOS logic may actually be smaller than the simple transistor count suggests.

Layout of Simple CMOS Logic Gates

In the following, we will examine simplified layout examples for CMOS NOR2 and NAND2 gates. Figure 7.14 shows a sample layout of a CMOS NOR2 gate, using single-layer metal and single-layer polysilicon.

In this example, the p-type diffusion area for pMOS transistors and the n-type diffusion area for nMOS transistors are aligned in parallel to allow simple routing of the gate signals via two parallel polysilicon lines running vertically. Figure 7.15 shows the layout of a CMOS NAND2 gate, using the same basic layout principles as in the NOR2 layout example.

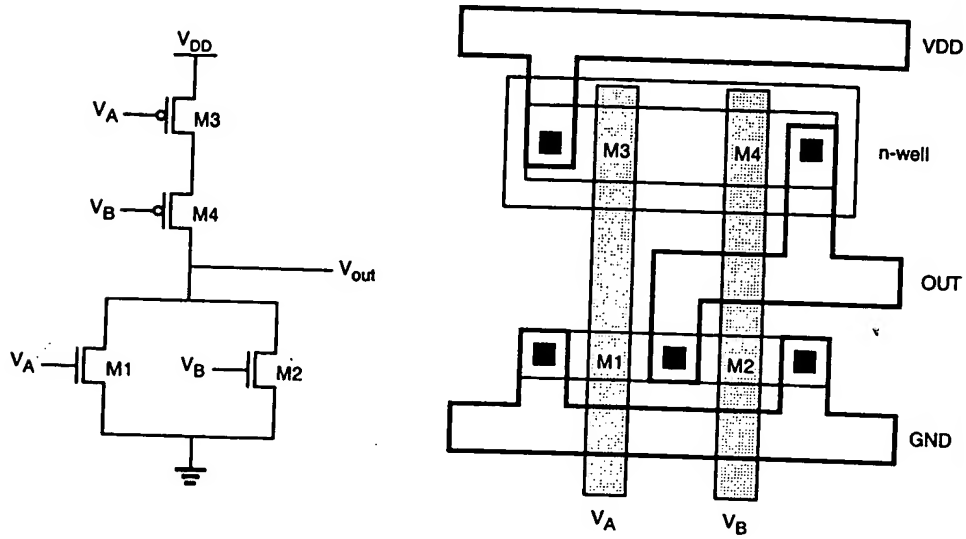


Figure 7.14. Sample layout of the CMOS NOR2 gate.

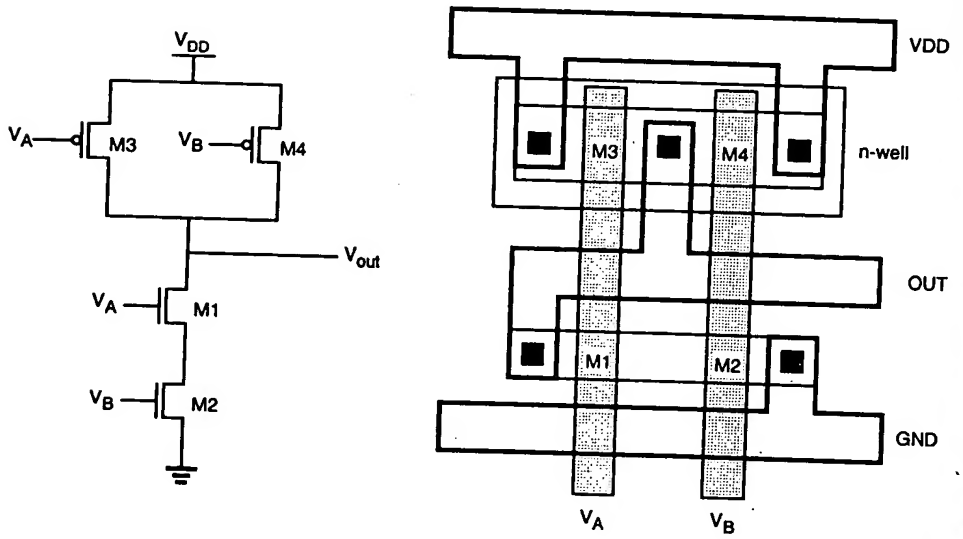


Figure 7.15. Sample layout of the CMOS NAND2 gate.

Finally, Fig. 7.16 shows a simplified (stick diagram) view of the CMOS NOR2 gate layout given in Fig. 7.14. Here, the diffusion areas are depicted by rectangles, the metal

connections and contacts are represented by solid lines and circles, respectively, and the polysilicon columns are represented by cross-hatched strips. The stick-diagram layout does not carry any information on the actual geometry relations of the individual features, but it conveys valuable information on the relative placement of the transistors and their interconnections.

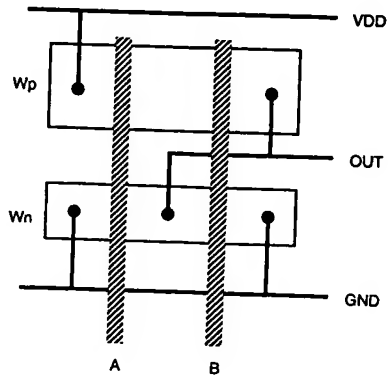


Figure 7.16. Stick-diagram layout of the CMOS NOR2 gate.

7.4. Complex Logic Circuits

To realize arbitrary Boolean functions of multiple input variables, the basic circuit structures and design principles developed for simple NOR and NAND gates in the previous sections can easily be extended to complex logic gates. The ability to realize complex logic functions using a small number of transistors is one of the most attractive features of nMOS and CMOS logic circuits.

Consider the following Boolean function as an example.

$$Z=\overline{A(D+E)+BC} \tag{7.41}$$

The nMOS depletion-load complex logic gate that is used to realize this function is shown in Fig. 7.17. Inspection of the circuit topology reveals the simple design principle of the pull-down network:

- OR operations are performed by parallel-connected drivers.
- AND operations are performed by series-connected drivers.
- Inversion is provided by the nature of MOS circuit operation.

The design principles stated here for individual inputs and corresponding driver transistors can also be extended to circuit sub-blocks, so that Boolean OR and AND operations can be performed in a nested circuit structure. Thus, we obtain a circuit topology which consists of series- and parallel-connected branches, as shown below.

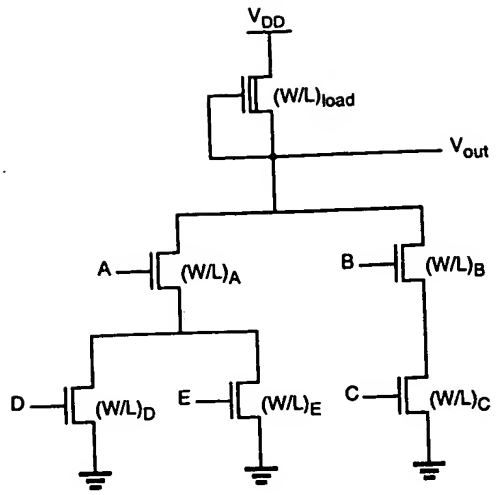


Figure 7.17. nMOS complex logic gate realizing the Boolean function given in (7.41).

In Fig. 7.17, the left nMOS driver branch consisting of three driver transistors is used to perform the logic function A(D + E), while the right-hand side branch performs the function BC. By connecting the two branches in parallel, and by placing the load transistor between the output node and the power supply voltage V_{DD}, we obtain the complex function given in (7.41). Each input variable is assigned to only one driver.

For the analysis and design of complex logic gates, we can employ the equivalent-inverter approach already used for the simpler NOR and NAND gates. It can be shown for the circuit in Fig. 7.17 that, if all input variables are logic-high, the equivalent-driver (*W/L*) ratio of the pull-down network consisting of five nMOS transistors is

$$\left(\frac{W}{L}\right)_{equivalent} = \frac{1}{\frac{1}{\left(\frac{W}{L}\right)_B} + \frac{1}{\left(\frac{W}{L}\right)_C}} + \frac{1}{\frac{1}{\left(\frac{W}{L}\right)_A} + \frac{1}{\left(\frac{W}{L}\right)_D + \left(\frac{W}{L}\right)_E}} \tag{7.42}$$

For calculating the logic-low voltage level V_{OL}, we have to consider various cases, since the value of V_{OL} actually depends on the number *and* the configuration of the conducting nMOS transistors in each case. All possible configurations are tabulated below. Each configuration is assigned a class number which reflects the total resistance of the current path from V_{out} node to ground

A - D	Class 1
A - E	Class 1
B - C	Class 1
A - D - E	Class 2

A - D - B - C	Class 3
A - E - B - C	Class 3
A - D - E - B - C	Class 4

Assuming that all driver transistors have the same (*W/L*) ratio, a Class 1 path such as (B-C) has the highest series resistance, followed by Class 2, Class 3, etc. Consequently, the logic-low voltage levels corresponding to each class have the following order, where each subscript numeral represents the class number.

$$V_{OL1} > V_{OL2} > V_{OL3} > V_{OL4} \tag{7.43}$$

The design of complex logic gates is based on the same ideas as the design of NOR and NAND gates. We usually start by specifying a maximum V_{OL} value. The design objective is to determine the driver and load transistor sizes so that the complex logic gate achieves the specified V_{OL} value even in the worst case. The given V_{OL} value first allows us to find the $(W/L)_{load}$ and $(W/L)_{driver}$ ratios for an equivalent inverter. Next, we have to identify all worst-case (Class 1) paths in the circuit, and determine the transistor sizes in these worst-case paths such that each Class 1 path has the equivalent driver ratio of $(W/L)_{driver}$.

In this example, this design strategy yields the following ratios for the three worst-case paths.

$$\begin{aligned}
\left(\frac{W}{L}\right)_A &= \left(\frac{W}{L}\right)_D = 2\left(\frac{W}{L}\right)_{driver} \\
\left(\frac{W}{L}\right)_A &= \left(\frac{W}{L}\right)_E = 2\left(\frac{W}{L}\right)_{driver} \\
\left(\frac{W}{L}\right)_B &= \left(\frac{W}{L}\right)_C = 2\left(\frac{W}{L}\right)_{driver}
\end{aligned} \tag{7.44}$$

The transistor sizes found above guarantee that, for all other input combinations, the logic-low output voltage level will be less than the specified V_{OL}.

Complex CMOS Logic Gates

The realization of the n-net, or pull-down network, is based on the same basic design principles examined earlier. The pMOS pull-up network, on the other hand, must be the dual network of the n-net. This means that all parallel connections in the nMOS pull-down network will correspond to a series connection in the pMOS pull-up network, and all series connections in the pull-down network correspond to a parallel connection in the pull-up network.

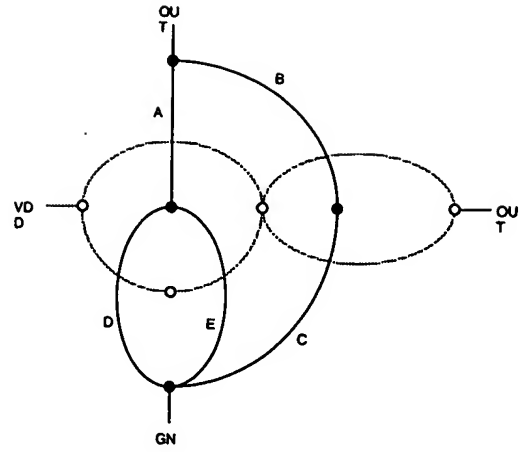


Figure 7.18. Construction of the dual pull-up graph from the pull-down graph, using the dual-graph concept.

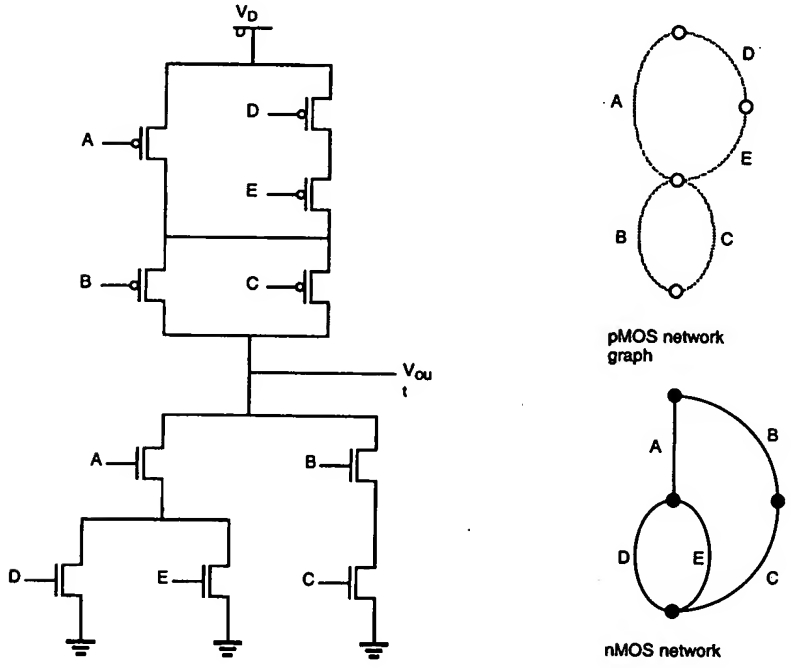


Figure 7.19. A complex CMOS logic gate realizing the Boolean function (7.41).

Figure 7.18 shows the simple construction of the dual p-net (pull-up) graph from the n-net (pull-down) graph. Each driver transistor in the pull-down network is represented by an edge, and each node is represented by a vertex in the pull-down graph. Next, a new vertex is created within each confined area in the pull-down graph, and neighboring vertices are connected by edges which cross each edge in the pull-down graph only once. This new graph represents the pull-up network. The resulting CMOS complex logic gate is shown in Fig. 7.19.

Layout of Complex CMOS Logic Gates

Now, we will investigate the problem of constructing a minimum-area layout for the complex CMOS logic gate. Figure 7.20 shows the stick-diagram layout of a "first attempt," using an arbitrary ordering of the polysilicon gate columns. Note that in this case, the separation between the polysilicon columns must allow for one diffusion-to-diffusion separation and two metal-to-diffusion contacts in between. This certainly consumes a considerable amount of extra silicon area.

If we can minimize the number of diffusion-area breaks both for nMOS and for pMOS transistors, the separation between the polysilicon gate columns can be made smaller, which will reduce the overall horizontal dimension and, hence, the circuit layout area. The number of diffusion breaks can be minimized by changing the *ordering* of the polysilicon columns.

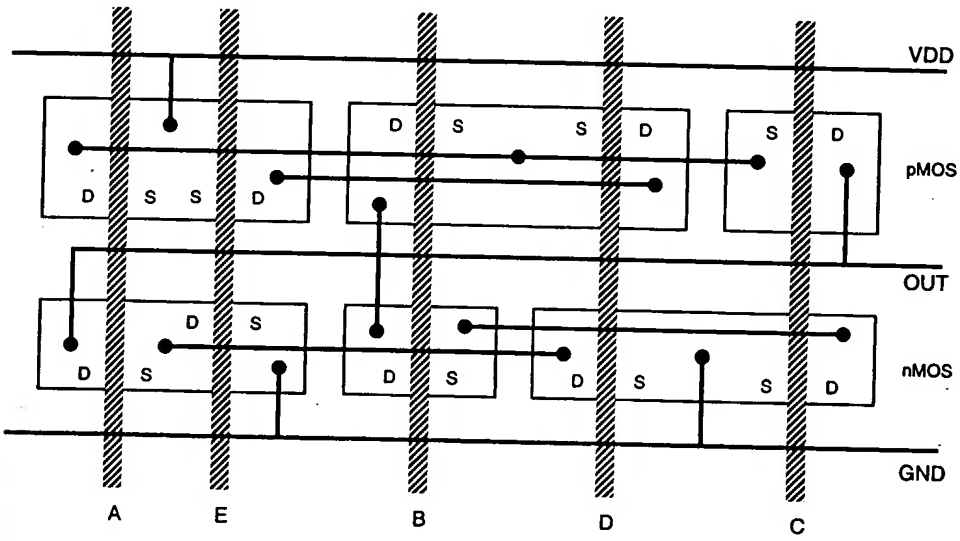


Figure 7.20. Stick-diagram layout of the complex CMOS logic gate, with an arbitrary ordering of the polysilicon gate columns.

A simple method for finding the optimum gate ordering is the Euler-path approach: find a Euler path in the pull-down graph and a Euler path in the pull-up graph with

identical ordering of input labels, i.e., find a common Euler path for both graphs. The Euler path is defined as an uninterrupted path that traverses each edge (branch) of the graph exactly once. Figure 7.21 shows the construction of a common Euler path for both graphs in our example.

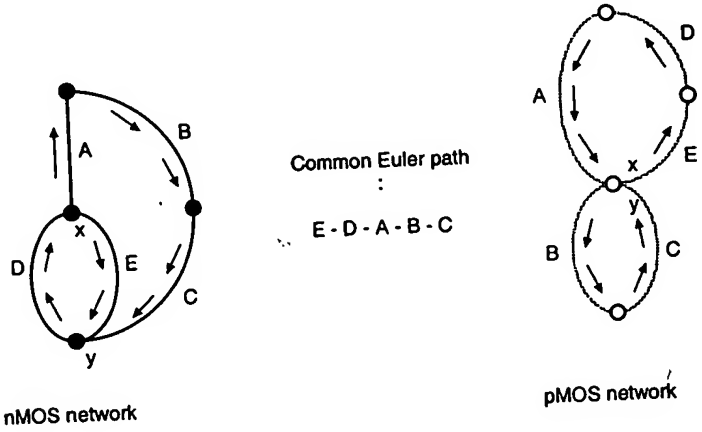


Figure 7.21. Finding a common Euler path in both graphs for n-net and p-net provides a gate ordering that minimizes the number of diffusion breaks and, thus, minimizes the logic- gate layout area. In both cases, the Euler path starts at (x) and ends at (y).

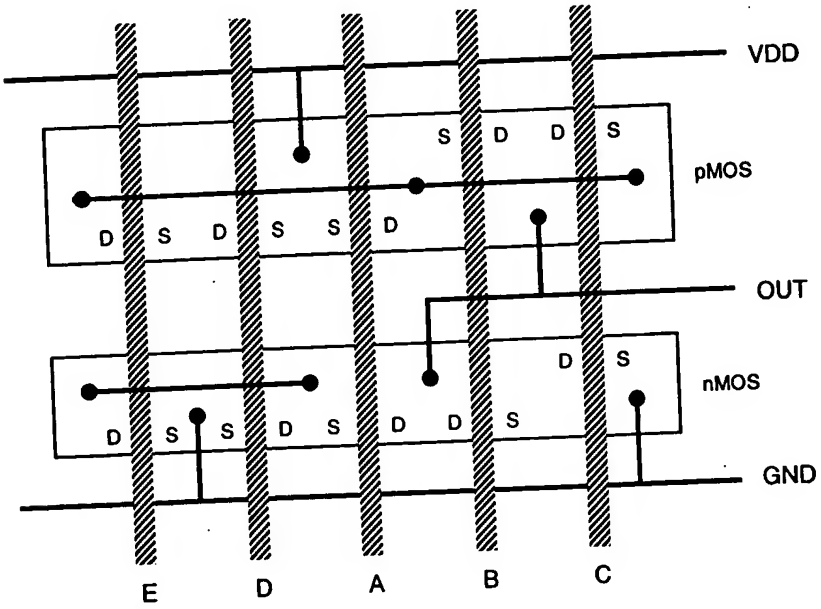


Figure 7.22. Optimized stick-diagram layout of the complex CMOS logic gate.

It is seen that there is a common sequence (E - D - A - B - C) in both graphs, i.e., a Euler path. The polysilicon gate columns can be arranged according to this sequence, which results in uninterrupted p-type and n-type diffusion areas. The stick diagram of the new layout is shown in Fig. 7.22. In this case, the polysilicon column separation Δd has to allow for only one metal-to-diffusion contact. The advantages of this new layout are more compact (smaller) layout area, simple routing of signals, and consequently, less parasitic capacitance.

As a further example of complex CMOS gates, the full-CMOS implementation of the exclusive-OR (XOR) function is shown in Fig. 7.23. Note that two additional inverters are also needed to obtain the inverse of both input variables (A and B). With these inverters, the CMOS XOR circuit in Fig. 7.23 requires a total of 12 transistors. Other CMOS realizations of the XOR gate that can be implemented with fewer transistors will be examined later.

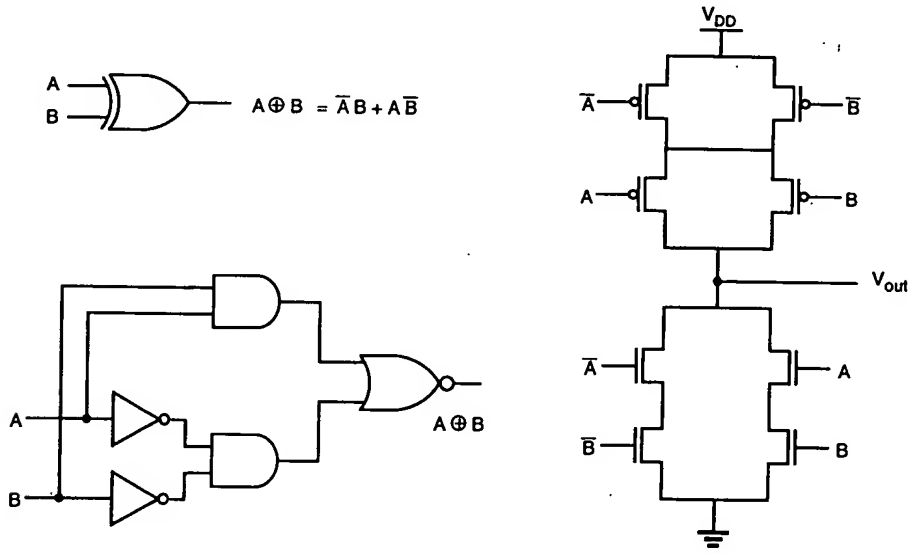


Figure 7.23. Full-CMOS implementation of the XOR function.

AOI and OAI Gates

While theoretically there are no strict limitations on the topology of the pull-down and the corresponding pull-up networks in a complex CMOS logic gate, we may recognize two important circuit categories as subsets of the general complex CMOS gate topology. These are the AND-OR-INVERT (AOI) gates and the OR-AND-INVERT (OAI) gates. The AOI gate, as its name suggests, enables the sum-of-products realization of a Boolean function in one logic stage (Fig. 7.24). The pull-down net of the AOI gate consists of parallel branches of series-connected nMOS driver transistors. The corresponding p-type pull-up network can simply be found using the dual-graph concept.

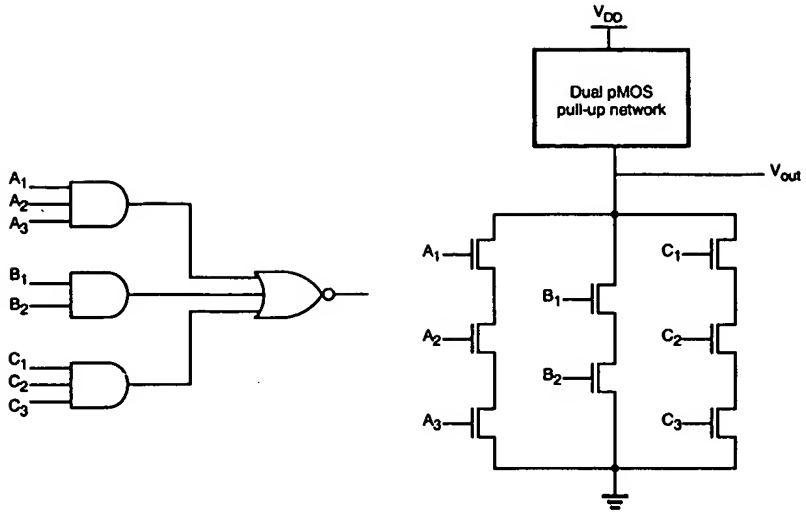


Figure 7.24. An AND-OR-INVERT (AOI) gate and the corresponding pull-down net.

The OAI gate, on the other hand, enables the product-of-sums realization of a Boolean function in one logic stage (Fig. 7.25). The pull-down net of the OAI gate consists of series branches of parallel-connected nMOS driver transistors, while the corresponding p-type pull-up network can be found using the dual-graph concept.

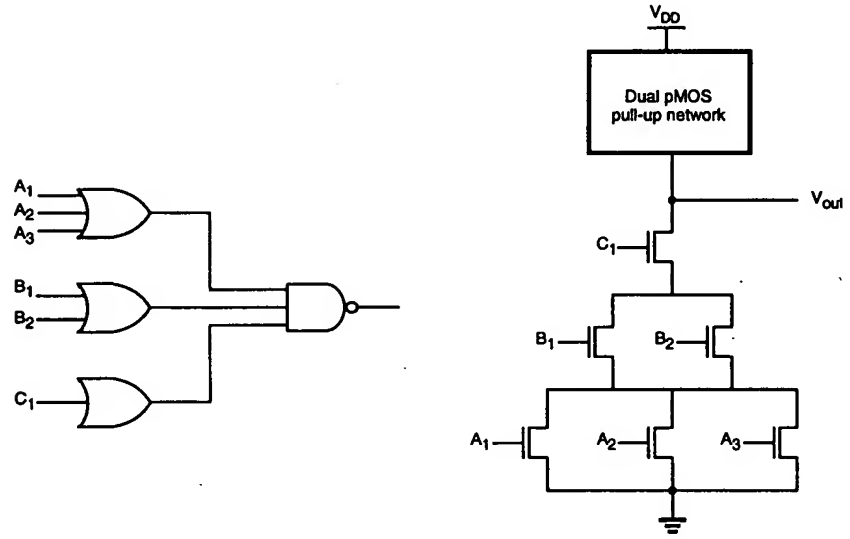


Figure 7.25. An OR-AND-INVERT (OAI) gate, and the corresponding pull-down net.

The large area requirements of complex CMOS gates present a problem in high-density designs, since two complementary transistors, one nMOS and one pMOS, are needed for every input. One possible approach to reduce the number of transistors is to use a single pMOS transistor, with its gate terminal connected to ground, as the *load* device (Fig. 7.26). With this simple pull-up arrangement, the complex gate can be implemented with much fewer transistors. The similarities of pseudo-nMOS gates to depletion-load nMOS logic gates are obvious.

The most significant disadvantage of using a pseudo-nMOS gate instead of a full-CMOS gate is the nonzero static power dissipation, since the always-on pMOS load device conducts a steady-state current when the output voltage is lower than V_{DD}. Also, the value of V_{OL} and the noise margins are now determined by the *ratio* of the pMOS load transconductance to the pull-down or driver transconductance.

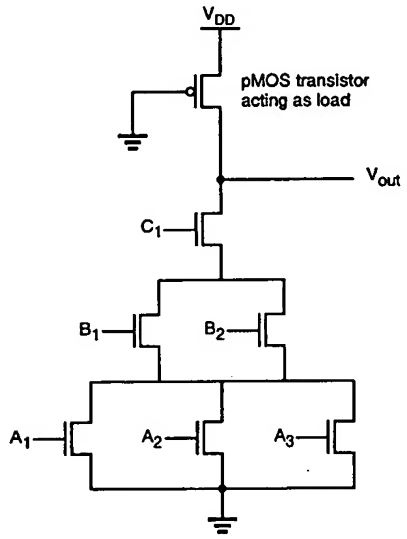
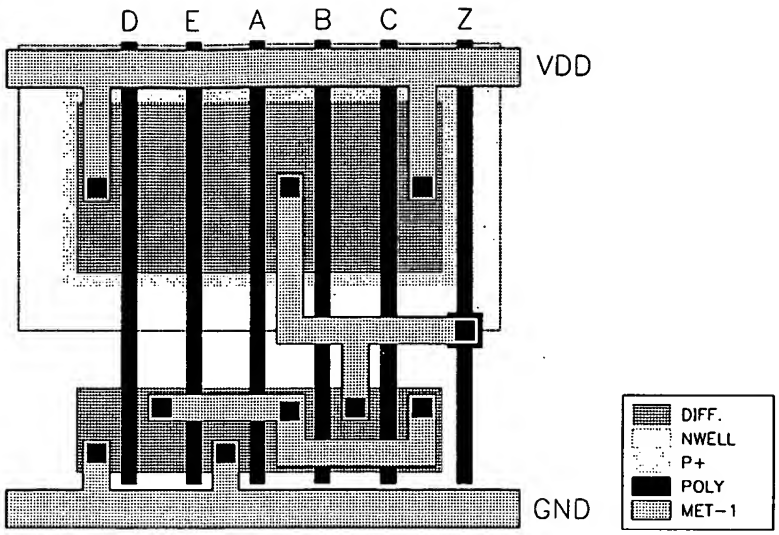


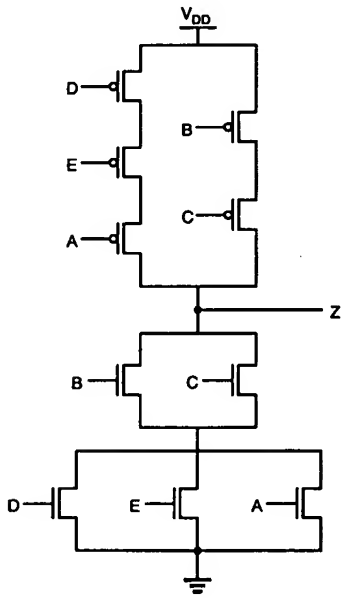
Figure 7.26. The pseudo-nMOS implementation of the OAI gate in Fig. 7.25.

Example 7.2.

The simplified layout of a CMOS complex logic circuit is given below. Draw the corresponding circuit diagram, and find an equivalent CMOS inverter circuit for simultaneous switching of all inputs, assuming that $(W/L)_p = 15$ for all pMOS transistors and $(W/L)_n = 10$ for all nMOS transistors.



The circuit diagram can be found from the layout by inspection:



The Boolean function realized by this circuit is

$$Z = \overline{(D+E+A)(B+C)}$$

The equivalent (*W/L*) ratios of the nMOS network and the pMOS network are determined by using the series-parallel equivalency rules discussed earlier in this chapter, as follows.

$$\left(\frac{W}{L}\right)_{n,eq} = \frac{1}{\frac{1}{\left(\frac{W}{L}\right)_D + \left(\frac{W}{L}\right)_E + \left(\frac{W}{L}\right)_A} + \frac{1}{\left(\frac{W}{L}\right)_B + \left(\frac{W}{L}\right)_C}}$$

$$= \frac{1}{\frac{1}{30} + \frac{1}{20}} = 12$$

$$\left(\frac{W}{L}\right)_{p,eq} = \frac{1}{\frac{1}{\left(\frac{W}{L}\right)_D} + \frac{1}{\left(\frac{W}{L}\right)_E} + \frac{1}{\left(\frac{W}{L}\right)_A}} + \frac{1}{\frac{1}{\left(\frac{W}{L}\right)_B} + \frac{1}{\left(\frac{W}{L}\right)_C}}$$

$$= \frac{1}{\frac{1}{15} + \frac{1}{15} + \frac{1}{15}} + \frac{1}{\frac{1}{15} + \frac{1}{15}} = 12.5$$

CMOS Full-Adder Circuit

The one-bit full adder circuit is one of the most widely used building blocks in all data processing (arithmetic) and digital signal processing architectures. In the following, we will examine the circuit structure and the realization of the full adder using the conventional CMOS design style.

The sum_out and carry_out signals of the full adder are defined as the following two combinational Boolean functions of the three input variables, *A*, *B*, and *C*.

$$\text{sum\_out} = A \oplus B \oplus C$$

$$= ABC + A\overline{B}\,\overline{C} + \overline{A}\,\overline{B}C + \overline{A}\,\overline{C}B$$

$$\text{carry\_out} = AB + AC + BC$$

A gate-level realization of these two functions is shown in Fig. 7.27. Note that instead of realizing the two functions independently, we use the carry_out signal to generate the sum output. This implementation will ultimately reduce the circuit complexity and, hence, save chip area. Also, we identify two separate sub-networks consisting of several gates (highlighted with dashed boxes) which will be utilized for the transistor-level realization of the full-adder circuit.

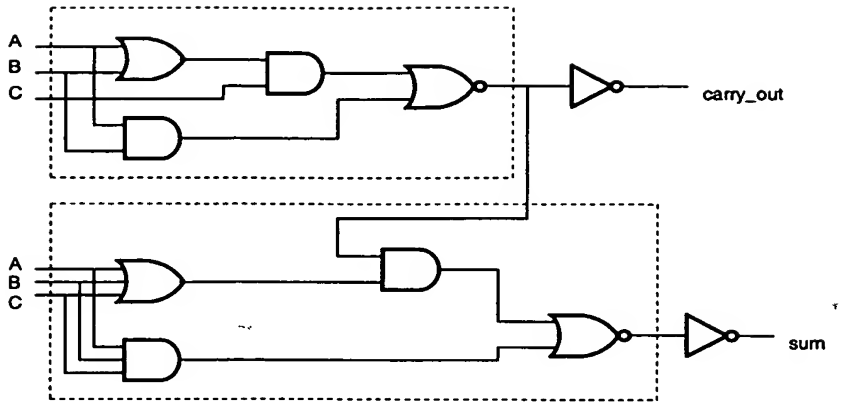


Figure 7.27. Gate-level schematic of the one-bit full-adder circuit.

The transistor-level design of the CMOS full-adder circuit is shown in Fig. 7.28. Note that the circuit contains a total of 14 nMOS and 14 pMOS transistors, together with the two CMOS inverters which are used to generate the outputs.

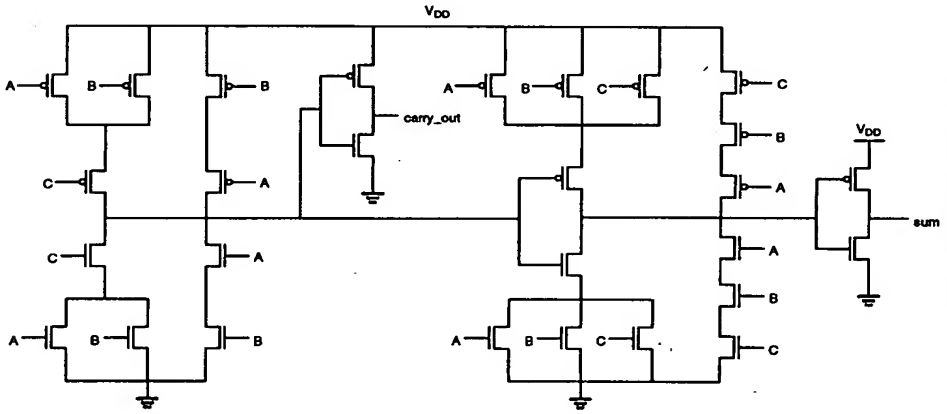


Figure 7.28. Transistor-level schematic of the one-bit full-adder circuit.

The mask layout of the full-adder circuit, which has been designed using the simple layout optimization strategy described earlier in this section, is shown in Fig. 7.29. Note, however, that all nMOS and pMOS transistors in this layout have the same (*W/L*) ratio. In order to optimize the transient (time-domain) performance of the circuit, it is usually necessary to adjust the transistor dimensions individually, as already shown in

Chapter 6. A performance-optimized and more compact mask layout of the same CMOS full-adder circuit is shown in Fig. 7.30.

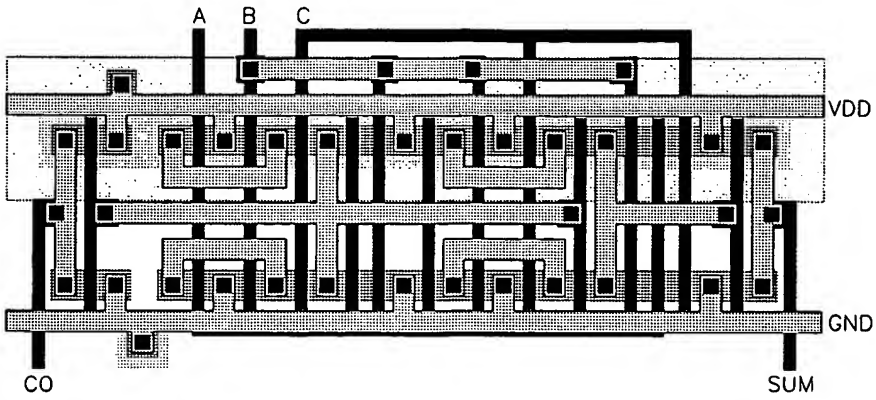


Figure 7.29. Mask layout of the CMOS full-adder circuit using minimum-size transistors.

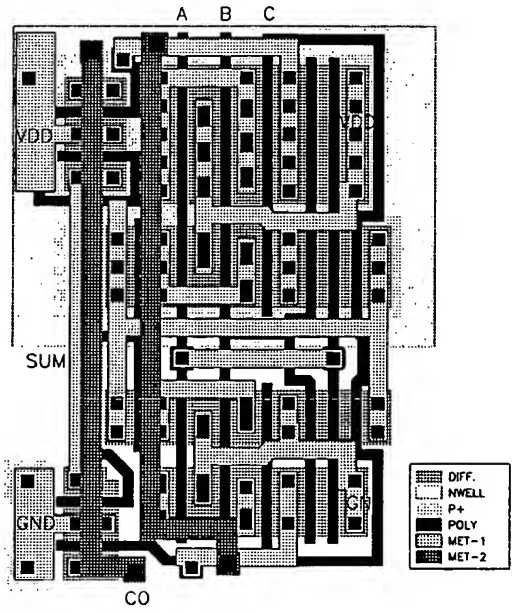


Figure 7.30. Mask layout of the optimized CMOS full adder circuit.

The simulated input and output voltage waveforms of the one-bit CMOS full adder are shown in Fig. 7.31. Please refer to the detailed design example that was presented in Chapter 1 for further design information.

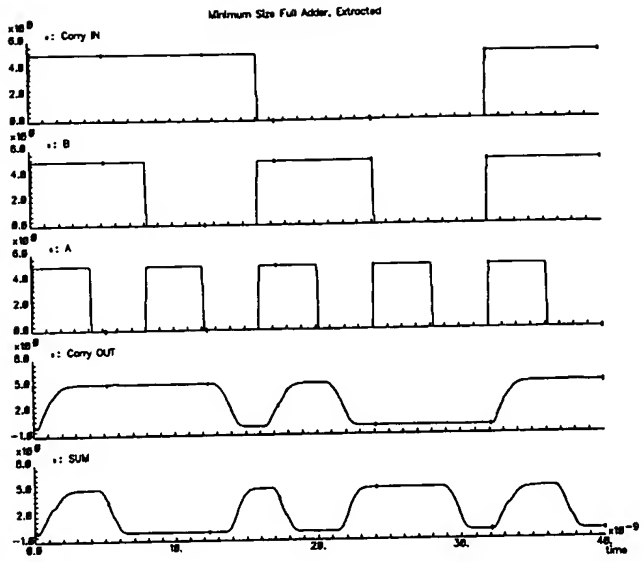


Figure 7.31. Simulated input and output waveforms of the CMOS full-adder circuit.

The full-adder circuit presented here can be used as the basic building block of a general n-bit binary adder, which accepts two n-bit binary numbers as input and produces the binary sum at the output. The simplest such adder can be constructed by a cascade-connection of full adders, where each adder stage performs a two-bit addition, produces the corresponding sum bit, and passes the carry output on to the next stage. Hence, this cascade-connected adder configuration is called the carry ripple adder (Fig. 7.32). The overall speed of the carry ripple adder is obviously limited by the delay of the carry bits rippling through the carry chain; therefore, a fast carry_out response becomes essential for the overall performance of the adder chain.

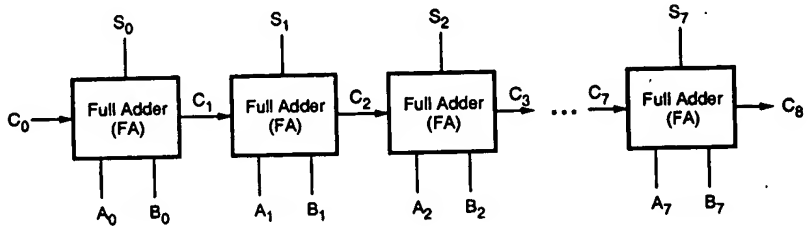


Figure 7.32. Block diagram of a carry ripple adder chain consisting of full adders.

7.5. CMOS Transmission Gates (Pass Gates)

In this section, we will examine a simple switch circuit called the CMOS transmission gate (TG) or pass gate, and present a new class of logic circuits which use the TGs as their basic building blocks. As shown in Fig. 7.33, the CMOS transmission gate consists of one nMOS and one pMOS transistor, connected in parallel. The gate voltages applied to these two transistors are also set to be complementary signals. As such, the CMOS TG operates as a bidirectional switch between the nodes A and B which is controlled by signal C.

If the control signal C is logic-high, i.e., equal to V_{DD}, then both transistors are turned on and provide a low-resistance current path between the nodes A and B. If, on the other hand, the control signal C is low, then both transistors will be off, and the path between the nodes A and B will be an open circuit. This condition is also called the high-impedance state.

Note that the substrate terminal of the nMOS transistor is connected to ground and the substrate terminal of the pMOS transistor is connected to V_{DD}. Thus, we must take into account the substrate-bias effect for both transistors, depending on the bias conditions. Figure 7.33 also shows three other commonly used symbolic representations of the CMOS transmission gate.

For a detailed DC analysis of the CMOS transmission gate, we will consider the following bias condition, shown in Fig. 7.34. The input node (A) is connected to a constant logic-high voltage, $V_{in} = V_{DD}$. The control signal is also logic-high, thus ensuring that both transistors are turned on. The output node (B) may be connected to a capacitor, which represents capacitive loading of the subsequent logic stages driven by the transmission gate. We will now investigate the input-output current-voltage relationship of the CMOS TG as a function of the output voltage V_{out}.

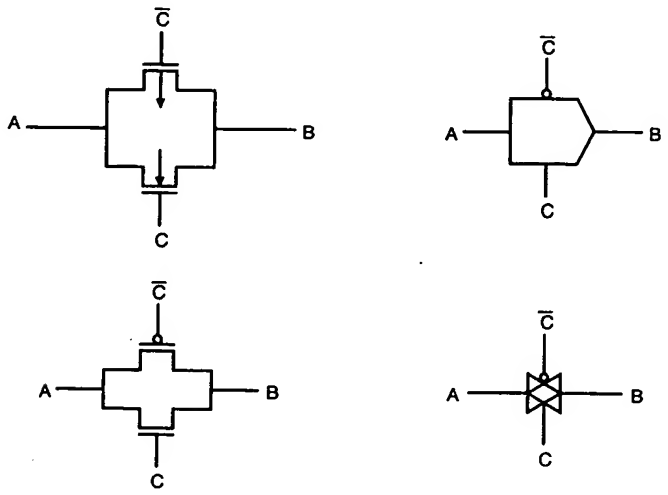


Figure 7.33. Four different representations of the CMOS transmission gate (TG).

It can be seen from Fig. 7.34 that the drain-to-source and the gate-to-source voltages of the nMOS transistor are

$$V_{DS,n} = V_{DD} - V_{out}$$
$$V_{GS,n} = V_{DD} - V_{out} \tag{7.45}$$

Thus, the nMOS transistor will be turned off for $V_{out} > V_{DD} - V_{T,n}$ and will operate in the saturation mode for $V_{out} < V_{DD} - V_{T,n}$. The V_{DS} and V_{GS} voltages of the pMOS transistor are

$$V_{DS,p} = V_{out} - V_{DD}$$
$$V_{GS,p} = -V_{DD} \tag{7.46}$$

Consequently, the pMOS transistor is in saturation for $V_{out} < |V_{T,p}|$, and it operates in the linear region for $V_{out} > |V_{T,p}|$. Note that, unlike the nMOS transistor, the pMOS transistor remains turned on, regardless of the output voltage level V_{out}.

This analysis has shown that we can identify three operating regions for the CMOS transmission gate, depending on the output voltage level. These operating regions are depicted in Fig. 7.34 as functions of V_{out}. The total current flowing through the transmission gate is the sum of the nMOS drain current and the pMOS drain current.

$$I_D = I_{DS,n} + I_{SD,p} \tag{7.47}$$

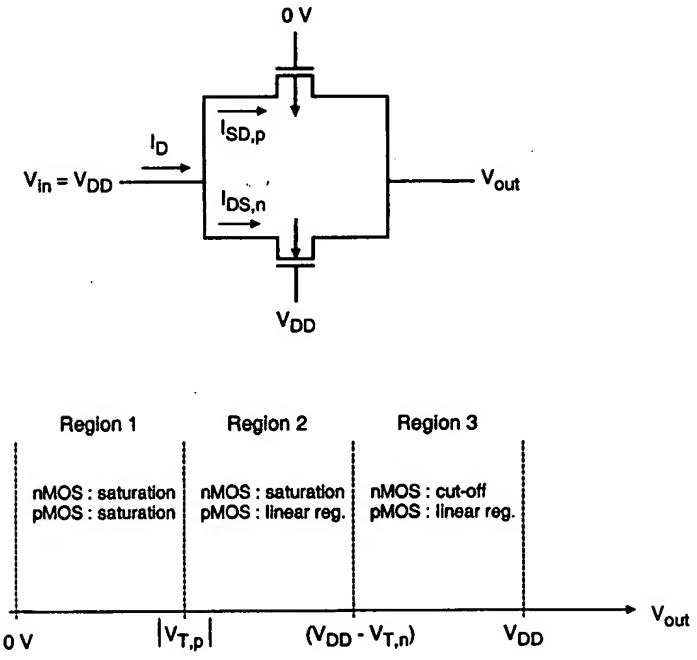


Figure 7.34. Bias conditions and operating regions of the CMOS transmission gate, shown as functions of the output voltage.

At this point, we may devise an *equivalent resistance* for each transistor in this structure, as follows.

$$R_{eq,n} = \frac{V_{DD} - V_{out}}{I_{DS,n}}$$
$$R_{eq,p} = \frac{V_{DD} - V_{out}}{I_{SD,p}} \tag{7.48}$$

The total equivalent resistance of the CMOS TG will then be the parallel equivalent of these two resistances, $R_{eq,n}$ and $R_{eq,p}$. Now, we will calculate the equivalent resistance values for the three operating regions of the transmission gate.

Region 1

Here, the output voltage is smaller than the absolute value of the pMOS transistor threshold voltage, i.e., $V_{out} < |V_{T,p}|$. According to Fig. 7.34, both transistors are in saturation. We obtain the equivalent resistance of both devices as

$$R_{eq,n} = \frac{2\left(V_{DD} - V_{out}\right)}{k_n\left(V_{DD} - V_{out} - V_{T,n}\right)^2} \tag{7.49}$$

$$R_{eq,p} = \frac{2\left(V_{DD} - V_{out}\right)}{k_p\left(V_{DD} - \left|V_{T,p}\right|\right)^2} \tag{7.50}$$

Note that the source-to-substrate voltage of the nMOS transistor is equal to the output voltage V_{out}, while the source-to-substrate voltage of the pMOS transistor is equal to zero. Thus, we have to take into account the substrate-bias effect for the nMOS transistor in our calculations.

Region 2

In this region, $|V_{T,p}| < V_{out} < (V_{DD} - V_{T,n})$. Thus, the pMOS transistor now operates in the linear region, while the nMOS transistor continues to operate in saturation.

$$R_{eq,n} = \frac{2\left(V_{DD} - V_{out}\right)}{k_n\left(V_{DD} - V_{out} - V_{T,n}\right)^2} \tag{7.51}$$

$$R_{eq,p} = \frac{2(V_{DD} - V_{out})}{k_p\left[2\left(V_{DD} - |V_{T,p}|\right)(V_{DD} - V_{out}) - (V_{DD} - V_{out})^2\right]}$$
$$= \frac{2}{k_p\left[2\left(V_{DD} - |V_{T,p}|\right) - (V_{DD} - V_{out})\right]} \quad (7.52)$$

Region 3

Here, the output voltage is $V_{out} > (V_{DD} - V_{T,n})$. Consequently, the nMOS transistor will be turned off, which results in an open-circuit equivalent. The pMOS transistor will continue to operate in the linear region.

$$R_{eq,p} = \frac{2}{k_p\left[2\left(V_{DD} - |V_{T,p}|\right) - (V_{DD} - V_{out})\right]} \quad (7.53)$$

Combining the equivalent resistance values found for the three operating regions, we can now plot the total resistance of the CMOS transmission gate as a function of the output voltage V_{out}, as shown in Fig. 7.35.

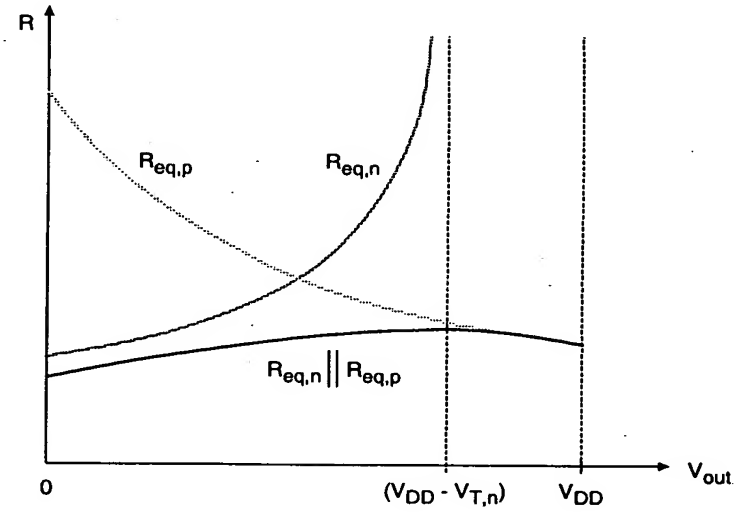


Figure 7.35. Equivalent resistance of the CMOS transmission gate, plotted as a function of the output voltage.

It can be seen that the total equivalent resistance of the TG remains relatively constant, i.e., its value is almost independent of the output voltage, whereas the individual equivalent resistances of both the nMOS and the pMOS transistors are strongly dependent on V_{out}. This property of the CMOS TG is naturally quite desirable. A CMOS pass gate which is turned on by a logic-high control signal can be replaced by its simple equivalent resistance for dynamic analysis, as shown in Fig. 7.36.

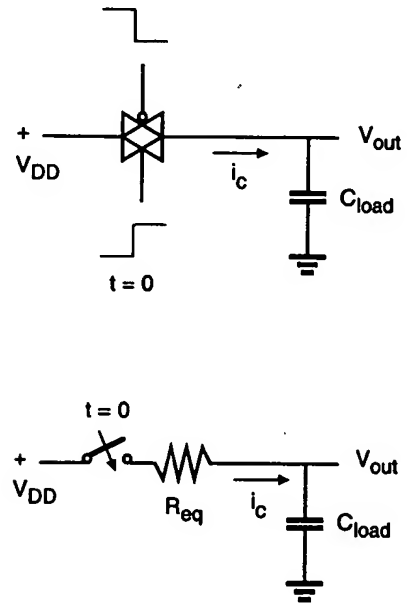


Figure 7.36. Replacing the CMOS TG with its resistor equivalent for transient analysis.

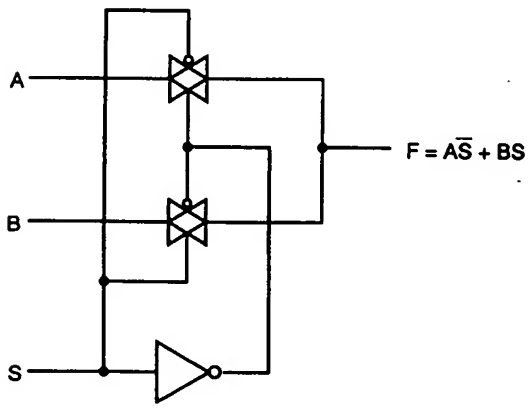


Figure 7.37. Two-input multiplexor circuit implemented using two CMOS TGs.

The implementation of CMOS transmission gates in logic circuit design usually results in compact circuit structures which may even require a smaller number of transistors than their standard CMOS counterparts. Note that the control signal *and* its complement must be available simultaneously for TG applications. Figure 7.37 shows a two-input multiplexor circuit consisting of two CMOS transmission gates. The operation of the multiplexor can be understood quite easily: If the control input S is logic-high, then the bottom TG will conduct, and the output will be equal to the input B. If the control signal is low, the bottom TG will turn off and the top TG will connect the input A to the output node.

Figure 7.38 shows an eight-transistor implementation of the logic XOR function, using two CMOS TGs and two CMOS inverters. The same function can also be implemented using only six transistors, as shown in Fig. 7.39.

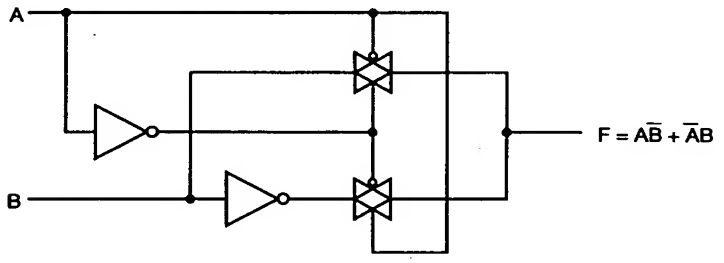


Figure 7.38. Eight-transistor CMOS TG implementation of the XOR function.

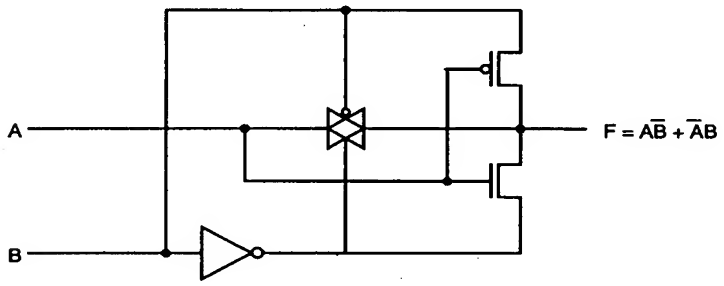
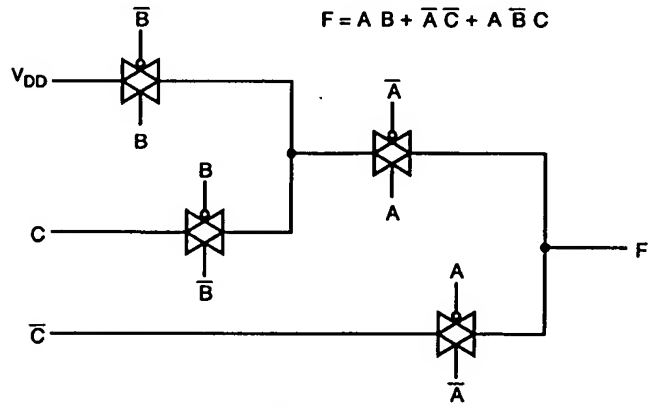


Figure 7.39. Six-transistor CMOS TG implementation of the XOR function.

Using the generalized multiplexor approach, each Boolean function can be realized with a TG logic circuit. As an example, Figure 7.40(a) shows the TG logic implementation of a three-variable Boolean function. Note that the three input variables *and* their inverses must be used to control the CMOS transmission gates. Including the three inverters not shown here, the TG implementation requires a total of 14 transistors. An

important point in TG logic design is that a conducting TG network (*low-impedance* path) should always be provided between the output node and one of the inputs, for all possible input combinations. This is to make sure that the output node with its capacitive load is never left in a *high-impedance* state.



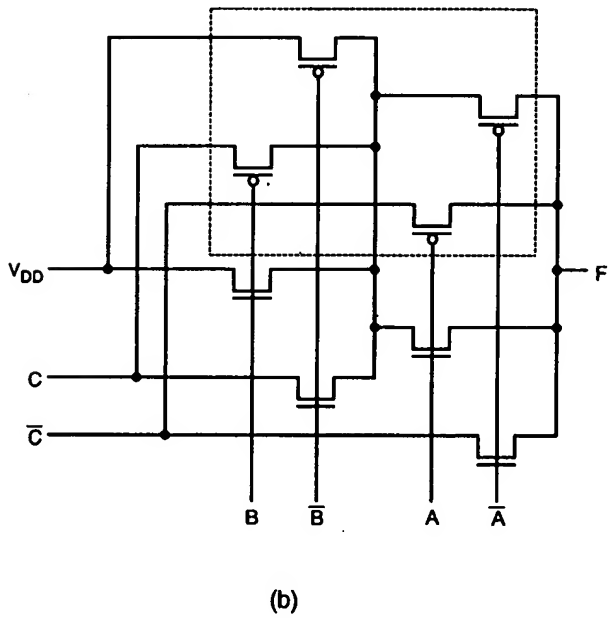


Figure 7.40. (a) CMOS TG realization of a three-variable Boolean function. (b) All pMOS transistors can be placed into one n-well to save area.

If each CMOS transmission gate in TG logic circuits is realized with a full nMOS-pMOS pair, the disjoint n-well structures of the pMOS transistors and the diffusion contacts may cause a significant overall area increase. In an attempt to reduce the silicon area occupied by TG circuits, the transmission gates can be laid out as separated nMOS-pMOS pairs with all pMOS transistors placed in one single n-well, as shown in Fig. 7.40(b). However, the *routing* area required for connecting the p-type diffusion regions to input signals must be carefully considered. The layout of the TG circuit is given in Fig. 7.41.

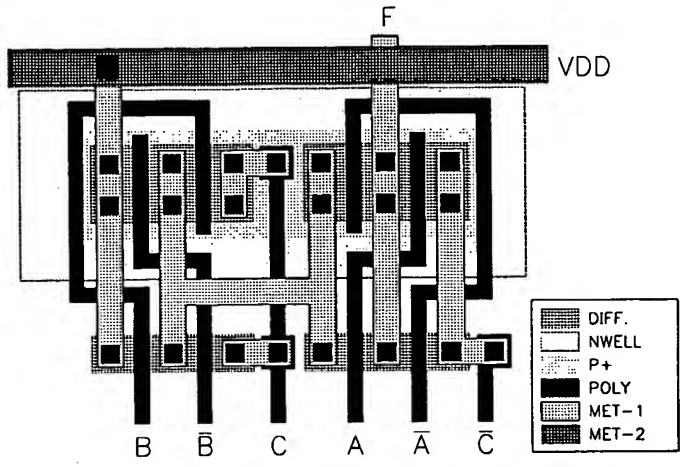


Figure 7.41. Mask layout of the CMOS TG circuit shown in Fig. 7.40.

Complementary Pass-Transistor Logic (CPL)

The complexity of full-CMOS pass-gate logic circuits can be reduced dramatically by adopting another circuit concept, called Complementary Pass-transistor Logic (CPL). The main idea behind CPL is to use a purely nMOS pass-transistor network for the logic operations, instead of a CMOS TG network. All inputs are applied in complementary form, i.e., every input signal and its inverse must be provided; the circuit also produces complementary outputs, to be used by subsequent CPL stages. Thus, the CPL circuit essentially consists of complementary inputs, an nMOS pass transistor logic network to generate complementary outputs, and CMOS output inverters to restore the output signals. The circuit diagrams of a CPL NOR2 and a CPL NAND2 are shown in Fig. 7.42.

The elimination of pMOS transistors from the pass-gate network significantly reduces the parasitic capacitances associated with each node in the circuit, thus, the operation speed is typically higher compared to a full-CMOS counterpart. But the improvement in transient characteristics comes at a price of increased process complexity. In CPL circuits, the threshold voltages of the nMOS transistors in the pass-gate network must be reduced to about 0 V through threshold-adjustment implants, in order

to eliminate the threshold-voltage drop. This, on the other hand, reduces the overall noise immunity and makes the transistors more susceptible to subthreshold conduction in the off-mode. Also note that the CPL design style is highly modular, a wide range of functions can be realized by using the same basic pass-transistor structures.

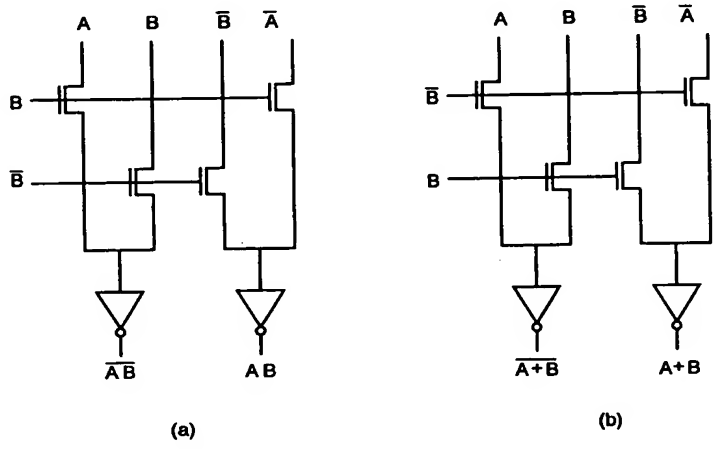


Figure 7.42. Circuit diagram of (a) CPL NAND2 gate and (b) CPL NOR2 gate.

Regarding the transistor count, CPL circuits do not always offer a marked advantage over conventional CMOS. The NAND2 and NOR2 circuits shown in Fig. 7.42 each consist of 8 transistors. XOR and XNOR functions realized with CPL have a similar complexity (i.e., transistor count) as conventional CMOS realizations. The same observation is true for the realization of full adders with CPL. The circuit diagram of a CPL-based XOR gate is shown in Fig. 7.43. Here, the cross-coupled pMOS pull-up transistors are used to speed up the output response. Transistor widths are given in λ-units. Figure 7.44 shows the circuit diagram of a CPL full-adder circuit consisting of 32 transistors. The mask layout of this CPL circuit is given in Fig. 7.45.

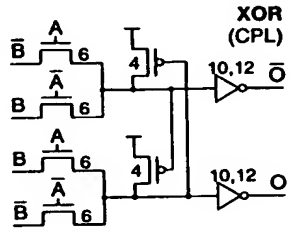


Figure 7.43. Circuit diagram of a CPL-based XOR gate.

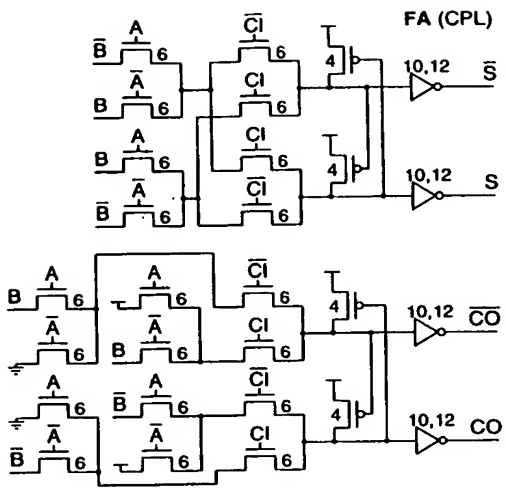


Figure 7.44. Circuit diagram of a CPL full adder.

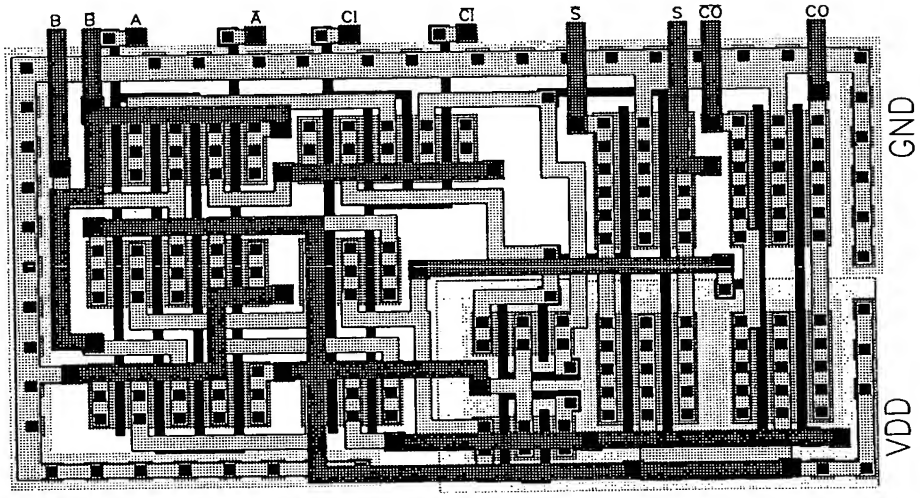


Figure 7.45. Mask layout of the CPL full adder shown in Fig. 7.44.

References

1. N.H.E. Weste and K. Eshraghian, *Principles of CMOS VLSI Design– A Systems Perspective*, second edition, Reading, MA: Addison-Wesley, 1993.

2. J.P. Uyemura, *Fundamentals of MOS Digital Integrated Circuits*, Reading, MA: Addison-Wesley, 1988.

3. L.A. Glasser and D.W. Dobberpuhl, *The Design and Analysis of VLSI Circuits*, Reading, MA: Addison-Wesley, 1985.

4. F.J. Hill and G.R. Peterson, *Computer-Aided Logical Design with Emphasis on VLSI*, fourth edition, New York, NY: John Wiley & Sons, Inc., 1993.

5. T. Sakurai and A.R. Newton, "Delay analysis of series-connected MOSFET circuits," *IEEE Journal of Solid-State Circuits*, vol. 26, no. 2, pp. 122-131, February 1991.

6. T. Uehara and W. M. Van Cleemput, "Optimal layout of CMOS functional arrays," *IEEE Transactions on Computers*, vol. C-30, no. 5, pp. 305-313, May 1981.

7. R. Zimmermann and W. Fichtner, "Low-power logic styles: CMOS versus pass-transistor logic," *IEEE Journal of Solid-State Circuits*, vol. 32, no. 7, pp. 1079-1090, July 1997.

GND

VDD

Exercise Problems

7.1 A CMOS circuit was designed based on company XYZ's 3-μm design rules, as shown in Fig. P7.1 with $W_N = 7$ μm and $W_P = 15$ μm.

(a) Determine the circuit configuration and draw the circuit diagram.
(b) For simple hand analysis, make the following assumptions:
i) Wiring parasitic capacitances and resistances are negligible.
ii) Device parameters are

	nMOS	pMOS
V_{T0}	1.0 V	−1.0 V
t_{ox}	500 Å	500 Å
k'	20 μA/V^2	10 μA/V^2
X_j	0.5 μm	0.5 μm
L_D	0.5 μm	0.5 μm

iii) The total capacitance at node I is 0.6 pF.
iv) An ideal step-pulse signal is applied to the CK terminal such that

$$V_{CK} = 5 \text{ V}, \quad t < 0$$
$$V_{CK} = 0 \text{ V}, \quad 0 \le t < T$$
$$V_{CK} = 5 \text{ V}, \quad t \ge T_w$$
$$V_{DD} = 5 \text{ V}$$

v) At $t = 0$, the node voltage at I is zero.
vi) The input voltages at A_1, B_1, and B_2 are zero for $0 \le t \le T_W$.

Find the minimum T_W that allows V_I to reach 2.5 V.

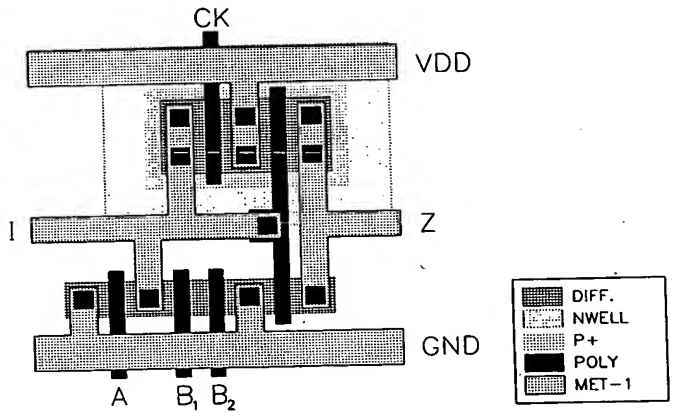


Figure P7.1

7.2 Calculate the equivalent W/L of the two nMOSTs with W_1/L and W_2/L connected in series. For simplicity, neglect the body effect, i.e., the threshold voltages of individual transistors are constant and do not depend on the source voltages. Although this is not true in reality, such an assumption is necessary for simple analysis with a reasonably good approximation.

7.3 Analytical expressions for V_{th}(logic) have been derived in Chapter 7 for the CMOS NOR2 gate. Now consider the CMOS NAND2 gate for the following cases and use $k_p = k_n = 100\ \mu A/V^2$:

- two inputs switching simultaneously
- top nMOS switching while the bottom nMOS's gate is tied to V_{DD}
- top nMOS gate is tied to V_{DD} and the gate input of the bottom nMOS is changing

(a) Derive an analytical expression for V_{th} corresponding to the first case. Also find the V_{th} value for the first case for $V_{DD} = 5$ V when the magnitudes of the threshold voltages are 1 V with $\gamma = 0$.

(b) Determine V_{th} for all three cases by using SPICE.

(c) For C_{load} = 0.2 pF, calculate 50% delays (low-to-high and high-to-low propagation delays) for an ideal pulse input signal for each of the three cases by assuming that C_{load} includes all of the internal parasitic capacitances. Verify the results using SPICE.

7.4 Write down the SPICE input description for transistor connections, source and drain parasitics in terms of areas, and perimeters for the layout shown in Example 7.2. Neglect the wiring capacitances in the polysilicon and metal runners. Default model names to be used for pMOS and nMOS are MODP and MODN. Assume L = 1 μm and $Y = 10$ μm for all transistors.

7.5 For the gate shown in Fig. P7.5,

- Pull-up transistor ratio is 5/5
- Pull-down transistor ratios are 100/5
- $V_{T0} = 1.0$ V
- $\gamma = 0.4\ V^{1/2}$
- $|2\phi_F| = 0.6$ V

(a) Identify the worst-case input combination(s) for V_{OL}.
(b) Calculate the worst-case value of V_{OL}. (Assume that all pull-down transistors have the same body bias and initially, that $V_{OL} = 5\%\ V_{DD}$.)

where C_i represents the parasitic capacitance associated with each node in the circuit (including the output node) and α_{Ti} represents the corresponding node transition factor associated with that node. Hence, the terms in the parentheses in (11.5) represent the total amount of *charge* which is drawn from the power supply during each switching event. While (11.5) is a more exact representation of the switching power dissipation in a CMOS logic gate, its evaluation may be relatively complicated. Therefore, we will mainly rely on (11.4) to express the switching power dissipation in CMOS logic circuits.

Observations on Switching Power Reduction

Expressions (11.4) and (11.5) derived above for the average switching power dissipation of CMOS logic gates suggest that we have several different means for reducing the power consumption. These measures include (i) reduction of the power supply voltage V_{DD}, (ii) reduction of the voltage swing in all nodes, (iii) reduction of the switching probability (transition factor) and (iv) reduction of the load capacitance. Note that the switching power dissipation is also a linear function of the clock frequency, yet simply reducing the frequency would significantly diminish the overall system performance. Thus, the reduction of clock frequency would be a viable option only in cases where the overall throughput of the system can be maintained by other means.

The reduction of power supply voltage is one of the most widely practiced measures for low-power design. While such reduction is usually very effective, several important issues must be addressed so that the system performance is not sacrificed. In particular, we need to consider that reducing the power supply voltage leads to an increase of delay. Also, the input and output signal levels of a low-voltage circuit or module should be made compatible with the peripheral circuitry, in order to maintain correct signal transmission.

The reduction of switching activity requires a detailed analysis of signal transition probabilities, and implementation of various circuit-level and system-level measures such as logic optimization, use of gated clock signals and prevention of glitches. Finally, the load capacitance can be reduced by using certain circuit design styles and by proper transistor sizing. These and other methods for the reduction of switching power dissipation will be examined in detail in the following Sections.

Short-Circuit Power Dissipation

The switching power dissipation examined above is purely due to the energy required to charge up the parasitic load capacitances in the circuit, and the switching power is independent of the rise and fall times of the input signals. Yet, if a CMOS inverter (or a logic gate) is driven with input voltage waveforms with finite rise and fall times, both the nMOS and the pMOS transistors in the circuit may conduct *simultaneously* for a short amount of time during switching, forming a direct current path between the power supply and the ground, as shown in Fig. 11.4.

The current component which passes through both the nMOS and the pMOS devices during switching does not contribute to the charging of the capacitances in the circuit, and hence, it is called the *short-circuit current* component. This component is especially prevalent if the output load capacitance is small, and/or if the input signal rise and fall times are large, as seen in Fig. 11.5. Here, the input/output voltage waveforms and the components of the current drawn from the power supply are illustrated for a symmetrical

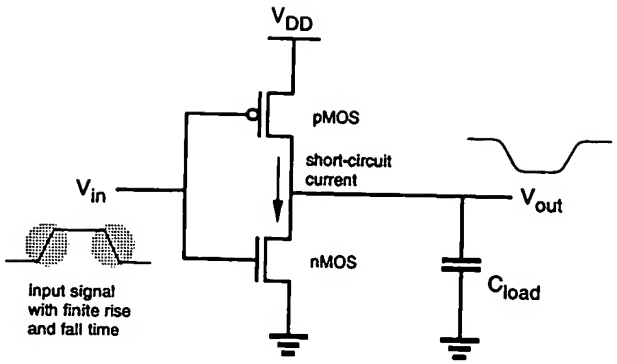


Figure 11.4. Both nMOS and pMOS transistor may conduct (simultaneously) a short-circuit current during switching.

CMOS inverter with small capacitive load. The nMOS transistor in the circuit starts conducting when the rising input voltage exceeds the threshold voltage $V_{T,n}$. The pMOS transistor remains on until the input reaches the voltage level (V_{DD} - $|V_{T,p}|$). Thus, there is a time window during which both transistors are turned on. As the output capacitance is discharged through the nMOS transistor, the output voltage starts to fall. The drain-to-source voltage drop of the pMOS transistor becomes nonzero, which allows the pMOS transistor to conduct as well. The short circuit current is terminated when the input voltage transition is completed and the pMOS transistor is turned off. A similar event is responsible for the short-circuit current component during the falling input transition, when the output voltage starts rising while both transistors are on.

Note that the magnitude of the short-circuit current component will be approximately the same during both the rising-input transition and the falling-input transition, assuming that the inverter is symmetrical and the input rise and fall times are identical. The pMOS transistor also conducts the current which is needed to charge up the small output load capacitance, but only during the falling-input transition (the output capacitance is discharged through the nMOS device during the rising-input transition). This current component, which is responsible for the switching power dissipation of the circuit (current component to charge up the load capacitance), is also shown in Fig. 11.5. The average of both of these current components determines the total amount of power drawn from the supply.

For a simple analysis consider a symmetric CMOS inverter with $k_n = k_p = k$ and $V_{T,n} = |V_{T,p}| = V_T$, and with a very small capacitive load. If the inverter is driven with an input voltage waveform with equal rise and fall times ($\tau_{rise} = \tau_{fall} = \tau$), it can be derived that the time-averaged short circuit current drawn from the power supply is

$$I_{avg}(short-circuit) = \frac{1}{12} \cdot \frac{k \cdot \tau \cdot f_{CLK}}{V_{DD}} (V_{DD} - 2V_T)^3 \qquad (11.6)$$

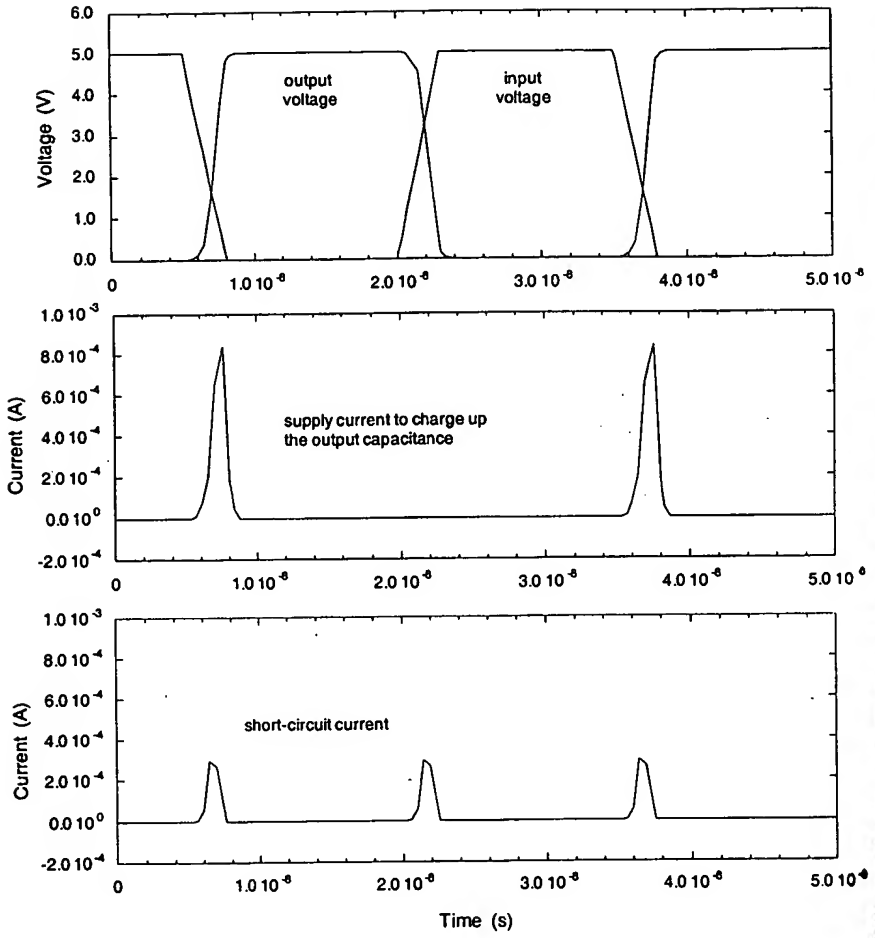


Figure 11.5. Input-output voltage waveforms, the supply current used to charge up the capacitance and the short-circuit current in a CMOS inverter with small capacitive load. The current drawn from the power supply is the sum of both current components.

Hence, the short-circuit power dissipation becomes

$$P_{avg}(short-circuit) = \frac{1}{12} \cdot k \cdot \tau \cdot f_{CLK} \cdot (V_{DD} - 2V_T)^3$$

Note that the short-circuit power dissipation is linearly proportional to the input rise and fall times, and also to the transconductance of the transistors. Hence, reducing the input transition times will decrease the short-circuit current component.

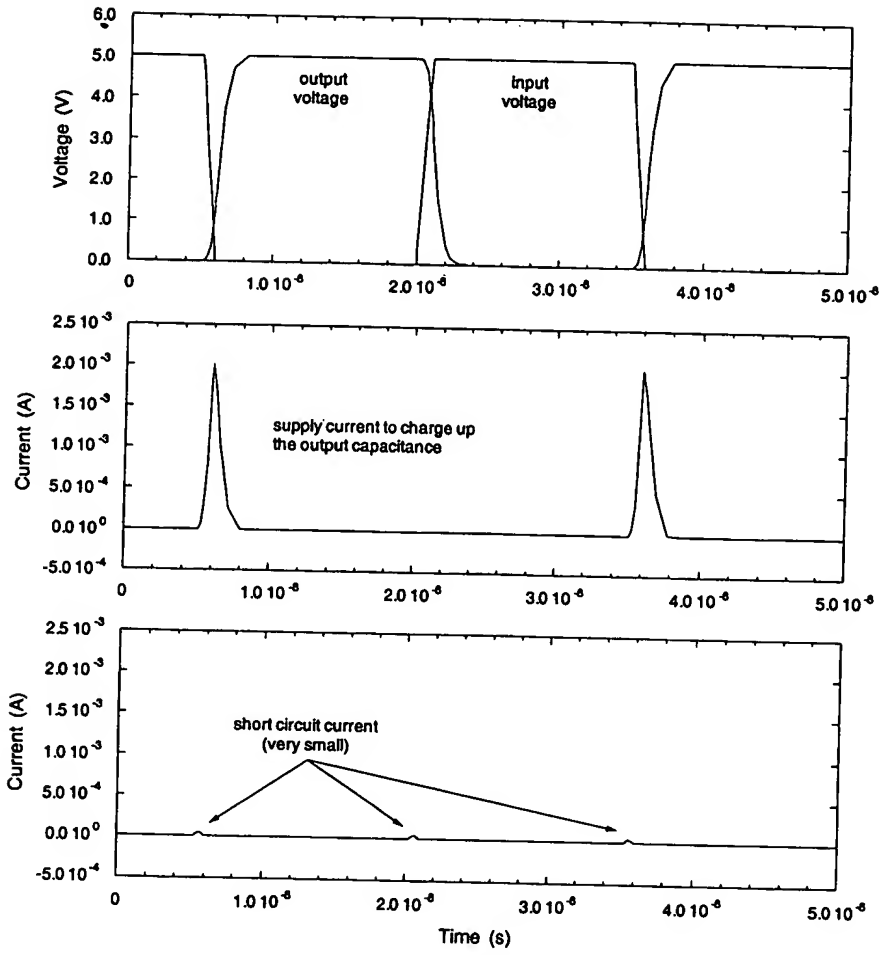


Figure 11.6. Input-output voltage waveforms, the supply current used to charge up the load capacitance and the short-circuit current in a CMOS inverter with larger capacitive load and smaller input transition times. The total current drawn from the power supply is approximately equal to the charge-up current.

Now consider the same CMOS inverter with a larger output load capacitance and smaller input transition times. During the rising input transition, the output voltage will effectively remain at V_{DD} until the input voltage completes its swing and the output will start to drop only after the input has reached its final value. Although both nMOS and pMOS transistors are on simultaneously during the transition, the pMOS transistor cannot conduct a significant amount of current since the voltage drop between its source and drain terminals is almost zero. Similarly, the output voltage will remain at approxi-

mately 0 V during the falling input transition and it will start to rise only after the input voltage completes its swing. Again, both transistors will be on simultaneously during the input voltage transition, yet the nMOS transistor will not be able to conduct a significant amount of current since its drain-to-source voltage is approximately zero. This situation is illustrated in Fig. 11.6, which shows the simulated input and output voltage waveforms of the inverter as well as the short-circuit and dynamic current components drawn from the power supply. Notice that the peak value of the supply current to charge up the output load capacitance is larger in this case. The reason for this is that the pMOS transistor remains in saturation during the entire input transition, as opposed to the previous case shown in Fig. 11.5 where the transistor leaves the saturation region before the input transition is completed.

The discussion concerning the magnitude of the short-circuit current may suggest that the short-circuit power dissipation can be reduced by making the output voltage transition times larger and/or by making the input voltage transition times smaller. Yet this goal should be balanced carefully against other performance goals such as propagation delay, and the reduction of the short-circuit current should be considered as one of the many design requirements that must be satisfied by the designer.

Leakage Power Dissipation

The nMOS and pMOS transistors used in a CMOS logic gate generally have nonzero reverse leakage and subthreshold currents. In a CMOS VLSI chip containing a very large number of transistors, these currents can contribute to the overall power dissipation even when the transistors are not undergoing any switching event. The magnitude of these leakage currents is determined mainly by the processing parameters.

Of the two main leakage current components in a MOSFET, the reverse diode leakage occurs when the pn-junction between the drain and the bulk of the transistor is revers-biased. The reverse-biased drain junction then conducts a reverse saturation current which is drawn from the power supply. Consider a CMOS inverter with a high input voltage, where the nMOS transistor is turned on and the output node voltage is discharged to zero. Although the pMOS transistor is turned off, there will be a reverse potential difference of V_{DD} between its drain and the n-well, causing a diode leakage through the drain junction. The n-well region of the pMOS transistor is also reverse biased with V_{DD}, with respect to the p-type substrate. Therefore, another significant leakage current component exists because of the n-well junction (Fig. 11.7).

A similar situation can be observed when the input voltage is zero, and the output voltage is charged up to V_{DD} through the pMOS transistor. Then, the reverse potential difference between the nMOS drain region and the p-type substrate causes a reverse leakage current which is also drawn from the power supply (through the pMOS transistor).

The reverse leakage current of a pn-junction is expressed by

$$I_{reverse} = A \cdot J_S \left(e^{\frac{q V_{bias}}{kT}} - 1 \right) \qquad (11$$